



COMILLAS

UNIVERSIDAD PONTIFICIA

ICAI

ICADE

CIHS

Facultad de Ciencias Humanas y Sociales
Grado en Traducción e Interpretación

Trabajo Fin de Grado

Linguistic Ambiguity:

Comparing Ambiguity in Standard
Mandarin Chinese and Spanish

Estudiante: María Irigoyen Ciriza

Directoras: Dra. Birgit Strotmann y Dra. Gisela Isabel Delfino López

Madrid, mayo 2023

ABSTRACT

Each language has its own kind of linguistic ambiguity, as its vocabulary and structures manifest it naturally based on the speaker's use of the language in a series of contexts. Linguistic ambiguity is a quality of language that leaves it open to interpretation. This paper looks at the three main types of linguistic ambiguity: lexical, semantic, and structural. For this, Standard Mandarin Chinese and Spanish have been chosen, since both languages manifest linguistic ambiguity in different forms. The paper also uses cognitive linguistics, more specifically cognitive grammar, to analyze how Standard Mandarin Chinese and Spanish solve linguistic ambiguity in some cases. This is done by providing examples of both languages corresponding to each kind of ambiguity.

Keywords: Cognitive linguistics, Standard Mandarin Chinese, Chinese, Spanish, culture, ambiguity.

RESUMEN

Cada lengua tiene su propio tipo de ambigüedad lingüística, ya que su vocabulario y sus estructuras la manifiestan de diferente manera a partir del uso que los hablantes hacen de la lengua en una serie de contextos. La ambigüedad lingüística es una cualidad de la lengua que hace que esta se pueda interpretar de dos o más maneras diferentes. Este trabajo examina los tres tipos principales de ambigüedad lingüística: léxica, semántica y estructural. Para ello, se han elegido el chino mandarín estándar y el español, ya que ambas lenguas manifiestan la ambigüedad lingüística de diferente forma. Además, el trabajo toma la lingüística cognitiva, en concreto la gramática cognitiva, como punto de partida para analizar cómo estas dos lenguas resuelven la ambigüedad lingüística en algunos casos. Para ello, se ofrecen ejemplos correspondientes a cada tipo de ambigüedad de ambas lenguas.

Palabras clave: lingüística cognitiva, chino mandarín estándar, chino, español, cultura, gramática, ambigüedad.

TABLE OF CONTENTS

<i>I. INTRODUCTION</i>	5
1. Purpose and Reason for Research	6
2. Structure	6
<i>II. THEORETICAL FRAMEWORK</i>	8
1. Mandarin	8
1.1. History.....	9
1.2. Relevant characteristics	10
1.3. Worldwide distribution	12
2. Spanish	13
2.1. History.....	13
2.2. Relevant characteristics	14
2.3. Worldwide distribution	16
3. Cognitive linguistics	17
3.1. Association.....	18
3.2. Automatization.....	19
3.3. Construal	20
3.4. Categorization	22
<i>III. METHODOLOGY</i>	24
<i>IV. ANALYSIS</i>	25
1. Lexical Ambiguity	25
2. Semantic Ambiguity	29
3. Structural Ambiguity	32
<i>V. SUMMARY AND CONCLUSION</i>	34
<i>VI. BIBLIOGRAPHY</i>	36
<i>VII. ANNEX</i>	39

TABLE OF VISUAL CONTENT

Figure 1.....	18
Figure 2.....	19
Figure 3.....	21
Figure 4.....	21
Table 1.....	11
Table 2.....	12
Table 3.....	15
Table 4.....	20
Table 5.....	22
Table 6.....	23
Table 7.....	25-26
Table 8.....	26
Table 9.....	27
Table 10.....	27-28
Table 11.....	29
Table 12.....	30
Table 13.....	31
Table 14.....	31

I. INTRODUCTION

When studying a language, be it from a native or learner perspective, people will encounter ambiguity. Each language has its own kind, as its vocabulary and structures present it naturally based on the speakers use of them in a series of contexts. Linguistic ambiguity is a quality of language that makes it open to interpretation. Something is ambiguous when the receiver can understand it in two or more possible ways. The concept of ambiguity is usually compared with vagueness: the former allows the creation of distinct interpretations, while the latter makes it difficult to interpret a message, as there is a lack of specificity. From a more formal point of view, there are three main types of linguistic ambiguity, these are: lexical, semantic, and structural ambiguity. Mandarin Chinese and Spanish are two languages that present linguistic ambiguity in different ways.

Standard Mandarin Chinese, or simply Mandarin, is the official language of China, officially the People's Republic of China (PRC). It is the variety of Chinese that is taught in schools and it is used in the working world. Based on Beijing's dialect, Mandarin is an artificial language based that was designed in the 20th century, as a measure to facilitate communication throughout the entire nation and to encourage the unification of its society. It is a language of non-alphabetical nature, consisting of characters that are developed logographs. As an estimation, there are more than 1.1 billion Mandarin Chinese speakers worldwide, of which around 900 million are native speakers.

Aside from being one of the languages with the highest numbers of speakers –Spanish has 496 million native speakers spread around the world–, Spanish is a language that does not share many characteristics with Mandarin. While it originated in the Iberian Peninsula, Spanish is now the official language of twenty countries and a dependent territory. This language developed from Vulgar Latin, which was the spoken language of the Roman Empire and now it is also known as Castilian in many areas. With these origins, Spanish uses the Latin or Roman script with the addition of an extra letter that is unique to the language. Moreover, it is thanks to King Alfonso X the Learned that Spanish spelling is fairly phonemic. Therefore, the pronunciation of a Spanish word can largely be predicted solely based on its spelling.

In the analysis section, both of these languages are examined in depth. Furthermore, this paper also covers how they solve linguistic ambiguity to an extent in certain cases. This is

explained by means of the cognitive abilities proposed by cognitive grammar, a point within cognitive linguistics.

Cognitive linguistics studies the relationship established between a language and the mind and, in fact, this line of research began by examining how language structure is influenced by outside factors. According to the International Cognitive Linguistics Association (n.d.), these factors include: structural characteristics of natural language categorization such as mental imagery, functional principles of linguistic organization, for instance, iconicity, and how language and thought are related, among other things. For this reason, it is important to also consider how cultures may or may not affect the way people communicate. In the case of Mandarin Chinese and Spanish, both cultures are highly dependent on context for different reasons, which can create a higher degree of ambiguity. Both Chinese and Spanish are world languages, and present a high level of linguistic ambiguity manifested in a different form, which is the reason why this paper will be comparing the two.

1. Purpose and Reason for Research

The study of ambiguity comparing Mandarin Chinese and Spanish is a field that is still fairly new. Considering the recent establishment of Standard Mandarin, the language is still only beginning to be studied, a process that is not particularly easy, seeing as languages develop at a rapid pace. Furthermore, Spanish is the official language of twenty countries, which makes studying it especially difficult. Although varieties are mutually intelligible, each country has its own way of speaking. Therefore, there may be different levels of ambiguity. For this reason, this paper will take to analyze the differences in ambiguity that Standard Mandarin Chinese and the Spanish variety of peninsular Spain present. Both languages have a high level of linguistic ambiguity, but they are not often compared together.

2. Structure

This paper will first tackle some topics that are essential for the full comprehension of the topic before the analysis. In this manner, the theoretical framework will first look at what cognitive linguistics is, focusing on cognitive grammar. Next, the paper will follow by

presenting Mandarin Chinese, starting with its origins and then the most relevant characteristics of the language; the same process will be carried out for the Spanish language. With this, the theoretical framework is finished, and the reader will now be able to fully comprehend the analysis. In it, this paper first tackles the three different types of linguistic ambiguity, and then compares both languages in each category. All of this is done in relation to cognitive linguistics.

This paper aims to compare both Standard Mandarin Chinese and Spanish when it comes to linguistic ambiguity from the point of view of cognitive grammar. Nevertheless, the length of the study is limited by its nature –an undergraduate final degree project–, which is why this line of research is left open to further investigation and could be expanded and substantiated in more depth.

II. THEORETICAL FRAMEWORK

This section will be subdivided into four main points. It will begin by tackling briefly the origins as well as the main characteristics of both Mandarin and Spanish. The author believes that having both of these things in mind is imperative; the main characteristics in particular as they are the formal part of the languages themselves. The theoretical framework finishes with a summary of the concept of cognitive linguistics as an introduction to the subject, but will focus on only cognitive grammar and its four functions.

1. Mandarin

Standard Mandarin Chinese is often referred to as *Putonghua* (普通话 – Pǔtōnghuà in pinyin), which literally means the common language. Although less popular, it is also known as *Guoyu* (国语 – guóyǔ in pinyin), which literally means the nation's language. This second term is mainly used in Taiwan and colloquially in Hong Kong; in mainland China, this word is used to refer to Taiwanese's variation of Standard Mandarin.

Standard Mandarin Chinese is the official language of China, officially the People's Republic of China (PRC). In order to make it less confusing, this paper will refer to Standard Mandarin Chinese as simply Mandarin. This variety of Chinese is the one that is taught in schools, as well as used in the working world. Mandarin is considered a sort of artificial language, as “it was deliberately designed in the early twentieth century to be distinct from any existing spoken vernacular” (Weng, 2018, p. 1). Based on Beijing's speech, Mandarin was created to facilitate communication throughout the entire nation, thus encouraging the unification of its society. Furthermore, another reason why the project that led the creation of this particular dialect was launched, was due to the difficulty of learning Classical Chinese, which was heavily restricted to a small segment of society. Therefore, Mandarin represented an attempt to extend educational meritocracy.

The word ‘Mandarin’ originates from “the Portuguese word *mandarim*, which developed, by way of Malay *měntěri*, from Sanskrit *mantrin*, meaning ‘counselor’” (Merriam-Webster Dictionary, n.d., para. 3). Eventually, the word evolved to carry the meaning of a bureaucrat, in other words, a person of position and influence. In the late 16th century, the English language

adopted ‘Mandarin’ as a noun, and it was not until the early 17th century that it began to be used as an adjective.

1.1. History

Mandarin belongs to the Sino-Tibetan language family, which directly descends from Proto-Sino-Tibetan (reconstructed proto-language). Eventually, it began to spread out and, thus evolved into a series of languages. In this manner, Old Chinese came to be. The earliest examples of Chinese date back to 1250 BCE and they were written by using the Oracle Bone Script.

A more recognizable form of Old Chinese, this one is also known as Classical or Literary Chinese (古文 - guwen), developed during the Zhou dynasty (1046 BCE-256 BCE), known for the establishment of certain political and cultural characteristics that would be identified with China for the next two millennia. Although it has been a subject of controversy, the overall consensus is that Chinese had not yet developed tones (Sagart, 2006). Classical Chinese has been used as the formal written language until the early 20th century. However, spoken Chinese did evolve into what is known as Middle Chinese (中古汉语 - zhongguhanyu); this is the ancestor of almost all Modern Chinese varieties.

Afterwards, Chinese was subjected to the development of a number of mutually intelligible dialects, which in turn gave way to different varieties of written language. During the Ming (1368-1644) and Qing (1644-1911/1912) dynasties, an official administrative language was created primarily based on the Old Mandarin dialect that was spoken in the southern capital, Nanjing. This was the first instance at which the language was referred to as ‘Mandarin’; now considered to be Middle Mandarin (官话 – guanhua).

At the beginning of the 20th century, a new official written language was introduced in order to replace Classical Chinese. However, it was later decided that the official language should be based on the Beijing dialect due to its importance. In this manner, Putonghua –what is now known as Standard Mandarin Chinese or, simply, Mandarin– was born.

For a more in-depth explanation of the history of Mandarin Chinese, refer to the first annex; this is a summarized version of it.

1.2. Relevant characteristics

Written Mandarin is undeniably a language of non-alphabetic nature as it consists of characters that are developed logographs –pictographs, ideographs, and phonetic compounds– to carry a more visual meaning. This makes reading unknown Mandarin words virtually impossible. However, Chinese characters tend to include a graphic element that will give an indication of the pronunciation of the character; the other part of the character is called ‘radical’, which, in most cases, it serves to provide a category to which a word will be related to (Sagart, 2006). Furthermore, pinyin is an alphabetic scheme that represents the latinized version of Mandarin words; its primary use is to serve as an auxiliary aid in the teaching of the language.

Chinese characters are written with strokes; there are eight main categories: dot (丶), hook (丨), horizontal (一), left-falling (丿), right-falling (㇇), rising (lower element of ㇇), turning (冫, ㇇, 乙, etc.), and vertical (丨). Furthermore, there are rules of stroke order do not always strictly apply to every situation and, therefore, are sometimes violated. In summary, horizontal strokes are written before vertical ones, as well as left-falling strokes before right-falling ones; characters are written from top to bottom and left to right; the middle is written first when the character is symmetric; closing a frame should be the last step (Björkstén, 1994).

As explained before, Mandarin is oftentimes referred to as an artificial language, seeing as, in China, people tend to speak in their own regional dialects and only speak Mandarin in school and other formal settings. Mandarin is also spoken when people belonging to different regions socialize, as it is the koiné language.

As well as many of the Eastern Asian languages, it is monosyllabic. Although there is no known language in which every morpheme is monosyllabic, this is generally taken to mean that the vast majority of them do consist of single syllables. In Mandarin, there are a great variety of polysyllabic words. However, they consist of a number of single syllables that, in most cases, carry their own meaning (Norman, 1988). An example of this would be the word ‘手机’, which means ‘mobile phone’ or ‘telephone’. It is clear that it consists of two characters, therefore, it is a polysyllabic morpheme. However, when taken apart, ‘手’ means ‘hand’ and ‘机’ means machine. The latter is used to refer to a large variety of machinery, technology, and transport; some examples: ‘耳机’ (literally ear machine) means ‘earphones’, and ‘飞机’ (literally flying machine) means ‘airplane’. This specific example will be used on the analysis.

Mandarin falls under the tonal language typology; they are characterized by a fixed pitch pattern according to level or contour. Norman (1988) highlights that tones are phonemic because they may serve to differentiate meaning like consonantal and vocalic segments. Mandarin has four tones:

- High-level represented as such: ‘ā’ or ‘a1’.
- Mid to high rising represented as such: ‘á’ or ‘a2’.
- Low falling-rising represented as such: ‘ǎ’ or ‘a3’.
- High to low falling represented as such: ‘à’ or ‘a4’.

Additionally, there is also a neutral tone, which it is sometimes not considered as a distinct tone because it is the absence of any kind of stress. Furthermore, it does not have a specific representation in pinyin. Below is a table that aims to showcase the different tones and the meanings they each carry, while, essentially, being the same word in pinyin.

妈	麻	马	骂	吗
mā	má	mǎ	mà	ma
Mom	Hemp	Horse	Insult/curse	<i>Particle used at the end of a question</i>

Table 1. Compiled by author.

Moreover, Mandarin uses classifiers, also known as measure words, with numerals and determinatives. The term ‘measure’ refers to what Y. R. Chao (1968) referred to as ‘individual measure’, which are words associated with particular nouns. These words’ objective is to classify nouns by type; ‘个’ (gè in pinyin) is versatile in nature as it can be used as a measure word for all nouns.

Furthermore, for the most part, Mandarin follows the SVO (Subject, Verb, and Object) word order. For instance:

S	V	O
我	说	中文。
wǒ	shuō	zhōngwén.
I	speak	Chinese.

Table 2. Compiled by author.

“Another important syntactic feature is the order of adjectives (and modifiers in general) and the nouns to which they refer. Chinese [...] place[s] the adjective before the noun” (Norman, 1988, p. 26).

Lastly, to mention some other details regarding punctuation:

- The period is a characteristic feature of the language: ‘ 。
- The comma is different to the one from languages that use the Latin script: ‘ ， ’

1.3. Worldwide distribution

Mandarin is the official language of China, Singapore, and Taiwan. Although not mentioned in this paper, Cantonese is a popular variety of Chinese and is the official language of Hong Kong, and Macau. It is difficult to give an accurate estimate of how many people from Hong Kong and Macau are fluent in Mandarin as the official education laws are fairly new. However, it seems that while not a big percentage of people are completely fluent, most of them do understand Mandarin to a certain extent.

Furthermore, in 1946, Mandarin Chinese became one of the six official languages of the United Nations. However, it was not until 1973-1974 that it became one of the working languages of the General Assembly in 1973, and the Security Council in 1974 (United Nations, n.d.).

According to a survey carried out by the Statista Research Department, as of 2022, there were “slightly more than 1.1 billion Mandarin Chinese speakers” (Statista Research

Department, 2023, para. 1); this number refers to fluent speakers only. It is estimated that around 900 million of them are native speakers, while the rest speak it as a second language.

2. Spanish

Spanish is a member of the Ibero Romance subfamily. In a similar way to many other Romance languages, Spanish developed from Vulgar Latin, which was the spoken language of the Roman Empire. Castilian is the official name of the Spanish variety spoken in Spain, and it was the one that became one of the most important languages in the world. While it originated in the Iberian Peninsula, Spanish is now the official language of twenty countries and a dependent territory.

2.1. History

The Spanish language belongs to the Indo-European family of languages; the first ancestor can be dated back to approximately 5,000 years ago around the area of the Black Sea. Ultimately, those who spoke Indo-European started to spread in various directions and, gradually, varieties of it began to appear. In this case, there is a particular branch that can be traced as the origins of Spanish, and that is the Italic branch.

Latin became the most prominent member of the Italic branch by means of political and cultural pressure from Ancient Rome. As Rome was spreading out and conquering more territories, Latin also did due to Roman soldiers settling in those territories. However, these soldiers did not speak Classical Latin (i.e. literary Latin), but what is known as Vulgar Latin. Vulgar Latin was less standardized in its grammar and vocabulary and is, essentially, “a spoken form of non-Classical Latin from which originated the Romance group of languages” (Sala, n.d., para. 1).

In time, different dialects descending from Vulgar Latin did also emerge within the Iberian Peninsula, more specifically what is now known as Spain, and slowly began to become distinct languages; the most important ones are: Aragonese, Castilian, Catalan, Galician-Portuguese, and Portuguese. In particular, Castilian arose in the north central region, in the province of Castille. This dialect was standardized by Alfonso X the Learned as the official language.

Moreover, he established a spelling system that successfully specified vernacular pronunciation making it fairly phonemic.

“It is in the 16th century that the term [Spanish] is applied to the language of culture of Spain, and therefore becomes equivalent in the sense to [Castilian]” (Penny, 2004, p. 31). In this manner, Spanish was born and became a world language, while Castilian was reduced to the specific name of the variation of Spanish spoken in peninsular Spain.

For a more in-depth explanation of the history of Spanish, refer to the second annex; this is a summarized version of it.

2.2. Relevant characteristics

As a descendant of Latin, Spanish uses the Latin or Roman script like the vast majority of European languages do. Researchers have pointed out that Latin script derives from the Greek alphabet that was used in some parts of southern Italy (Magna Grecia). The Spanish alphabet has a total of 27 letters, one more than the original Latin script has; this is because Spanish has the letter ‘ñ’. Moreover, as previously stated, it is thanks to Alfonso X the Learned that spelling is fairly phonemic. Therefore, the pronunciation of a Spanish word can largely be predicted solely based on its spelling.

The use of the Latin script allows capital letters. However, capitalization in Spanish is scarce compared to other languages; names of people and places, the first word of a sentence, the first word of the title of an artwork, some abbreviations, etc.

Morphemes are parts of every word that can consist of one or more letters. With this, words are analyzed, seeing as morphemes are capable of conveying a lexical meaning, sometimes called lexemes, and a grammatical function. The latter has a set of rules that are universal for the most part:

- The letter ‘o’ at the end of a word indicates that the gender of said word is masculine.
- The letter ‘a’ at the end of a word indicates that the gender of said word is feminine.
- The letter ‘s’ at the end of a word indicates plural.
- The letter ‘n’ at the end of a conjugated verb indicates plural.
- Some morphemes serve as conjugation markers and tense indicators.

For instance, ‘libros’ is a masculine and plural word due to the morphemes ‘o’ and ‘s’.

The Spanish languages allow a great amount of structure variations, the last two in the table show the pro-dropping of the subject:

S		V		O	
El libro		gusta		a los alumnos.	
O		V		S	
[A los alumnos] Les		gusta		el libro	
S		O		V	
El libro		les		gusta.	
S	O		V	O	
El libro	les		gusta	a los alumnos.	
V			O		
Gusta			a los alumnos.		
O			V		
[A los alumnos] Les			gusta.		

Table 3. Compiled by author.

Furthermore, Spanish conjugates verbs, they undergo some form of inflection that falls under the following categories:

- Mood: indicative, subjunctive, or imperative.
- Tense: past, present, or future.
- Person and number:

- a) first, second, or third
- b) singular or plural
- Aspect: perfective or imperfective

Lastly, in Spanish, syllables are not pronounced with equal emphasis; the stressed syllable is called *sílaba tónica*, while the rest of the syllables pronounced with less intensity are called *sílabas átonas*. In addition, the placement of the accent is subject to a number of strict orthographic rules.

2.3. Worldwide distribution

The following information has been obtained from the Instituto Cervantes' 2022 yearbook.

Spanish is the official language of twenty countries and one dependent territory; this means that 595 million people speak Spanish. Of those, 496 million are native speakers, and 75 million people, albeit not native, are fluent. Furthermore, it is the second language with the most native speakers and the fourth overall most spoken in the world –the first is English, the second is Mandarin Chinese, and the third is Hindi.

Spanish became one of the six official languages of the United Nations in 1946; two years after that, it was when the General Assembly adopted Spanish as a working language. Lastly, in 1969, it also became the working language of the Security Council; at present, it is the third more used language in the UN. Furthermore, in 1986, Spanish was added as one of the official languages. Nowadays, out of the 24 official languages, Spanish is the second most used in the EU.

The Cervantes Institute estimates that by 2060, the United States of America will be the second country with the largest number of Spanish speakers, after Mexico.

A 7.9% of Internet users communicate in Spanish, thus making it the third most used online. However, on Wikipedia as well as social media such as Facebook, YouTube, Twitter, and LinkedIn, Spanish is the second most used language.

3. Cognitive linguistics

After comparing Mandarin and Spanish are like, it becomes important to provide a starting point from where linguistic ambiguity can be solved. In this manner, this chapter will tackle the cognitive abilities related to cognitive grammar.

The study of cognitive linguistics began to emerge in the 1970s due to a growing interest to understand the relation of language and mind from a different perspective. In fact, there was a prevailing tendency to explain linguistic patterns by appealing to the structural properties that were internal to and specific to language. This new line of research began by attempting to examine how language structure was heavily influenced by outside factors. According to the International Cognitive Linguistics Association, these factors include: structural characteristics of natural language categorization (e.g. mental imagery, cognitive models, etc.), functional principles of linguistic organization (e.g. iconicity), the conceptual interface between syntax and semantics, the experiential and pragmatic background of language-in-use, and how language and thought are related.

Cognitive linguistics gradually grew in popularity as it provided a new approach to study the human mind by means of understanding languages. At present, it is endowed with an international society –International Cognitive Linguistics Association (ICLA)– that organizes related conferences and has a journal –*Cognitive Linguistics*.

This paper will specifically focus on cognitive grammar, as it provides insight into how languages solve the problem of linguistic ambiguity.

To understand cognitive grammar, one must first attempt to define the term ‘conceptualization’, which resides in cognitive processing and is grounded in physical reality. In this case, it refers to novel conceptions, as well as fixed ones, for instance, kinesthetic experiences or the recognition of a specific social context; this constitutes a mental experience. In this manner, cognitive grammar allows to draw a number of conclusions: the first one being the extent to which an expression’s meaning is determined by outside factors; and the second one would be the extent to which imaginative abilities become active. Furthermore, in order for a word to be symbolized phonologically, a language only needs three kinds of structures: semantic, phonological, and symbolic. Semantic structures represent conceptualizations for linguistic purposes; phonological structures refer to sounds, gestures –in the case of sign languages-, and orthographic representations; and symbolic structures incorporate the other

two. In his book, *Cognitive grammar: A basic introduction* (2008), Langacker proposed focusing on four cognitive functions: association, automatization, construal –and its subdivisions: schematization, prominence, and perspective/perspectivization–, and categorization.

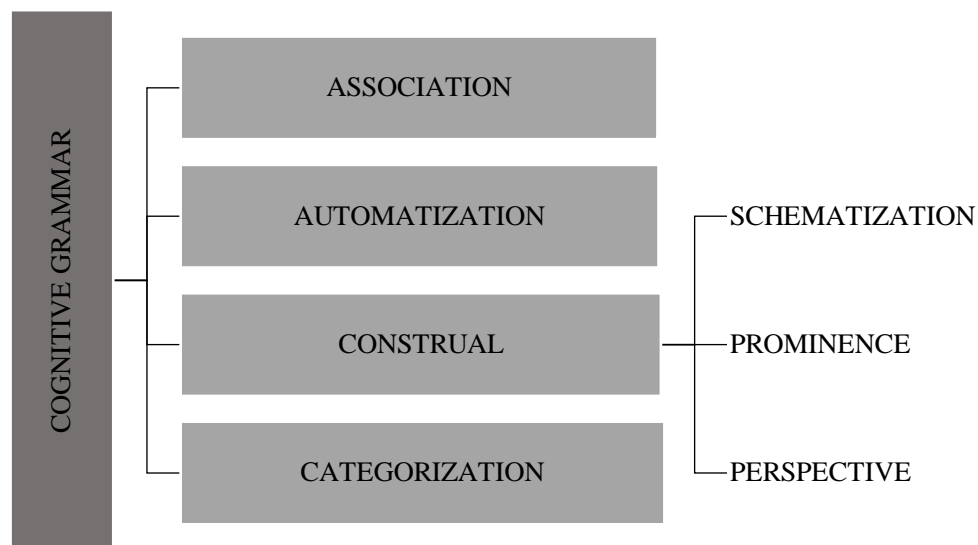


Figure 1. Source: Langacker R. W. (2008)

This table presents a summary of what this section of the paper will tackle in a more organized manner. It will be explained in that order.

3.1. Association

Association is the cognitive ability that focuses on establishing psychological connections that will have the potential to influence subsequent processing (Langacker, 2008). Association is a basic cognitive ability that establishes a connection between meanings and forms to create words and structures. In other words, this means that human minds pair a semantic pole and a phonological pole, the result of it is called a symbolic assembly. For instance, the lexical item ‘tree’:

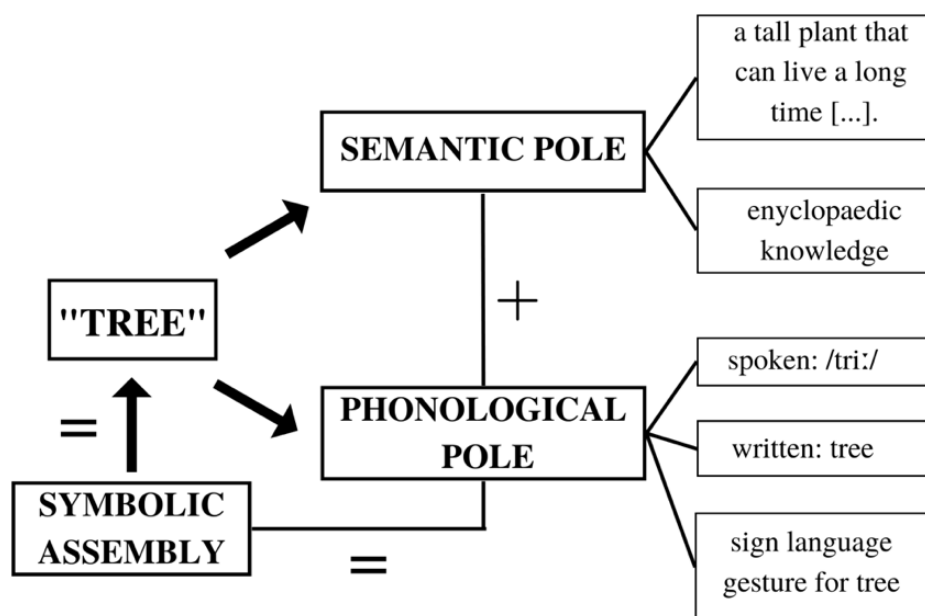


Figure 2. Compiled by author. Source: Langacker R. W. (2008).

In cognitive grammar, not only is a lexical item's *raison d'être* explained in this manner, but also grammar in its entirety. This means that a set of symbolic assemblies are combined to form complex syntactic structures. For this reason, Langacker points out that lexicon and syntax form a continuum.

3.2. Automatization

Automatization is a cognitive process that allows society to learn something by repeating it in order for it to become automatic. For instance, this can be walking home from school, tying a shoelace, etc. Cognitive grammar also illustrates how automatization is required to learn language structures –words, syntactic structures, etc.–; Langacker calls this process *entrenchment*. This is the process through which a specific structure undergoes entrenchment and to become established as a unit (Langacker, 2008). He gives the following example: 'penniless', which is an adjective that everybody knows and uses, therefore it is completely entrenched in the English language. However, the word 'dollarless', while it theoretically carries the same meaning, it is not entrenched in the language.

PENNY (lexical unit meaning ‘money’) + LESS (lexical unit) = PENNILESS

↳ Usage has made ‘penniless’ a single lexical unit; it is used as an adjective.

DOLLAR (lexical unit meaning ‘money’) + LESS (lexical unit) = DOLLARLESS

↳ It is not used; therefore, it is not a valid lexical unit.

Table 4. Compiled by author.

Langacker also reminds the reader that units must not be regarded as a list of items, but, rather, they should be considered to be a part of a complex network of relations.

3.3. Construal

Construal refers to a human’s ability to construe a situation from different approaches with various degrees of specificity. These are the different subdivisions of construal:

- Schematization: This subdivision forms conceptual delineations that can be ascribed to the pattern or to the quantity that is interrelated, whether in space or time or some form of conceptual dimension (Talmy, 2006). Essentially, schematization serves to provide a ‘dimension’ that encompasses all its generalizations, such as the ‘partitioning’ of time and space –specified as *this* and *that*. For this reason, schemas motivate aspects of how people imagine, reason, and think (Gibbs & Colston, 2006). Essentially, schemas can be understood as the general idea of something. An example could be the word ‘music’, which is a schematic unit, and is a general term that can, for instance, refer to ‘flamenco’, which would be a more specific term for ‘music’. In other words, ‘music’ is the dimension or schema, and ‘flamenco’ is a schematic unit that belongs to the dimension of music. The following graphic offers another example of schemas and dimensions:

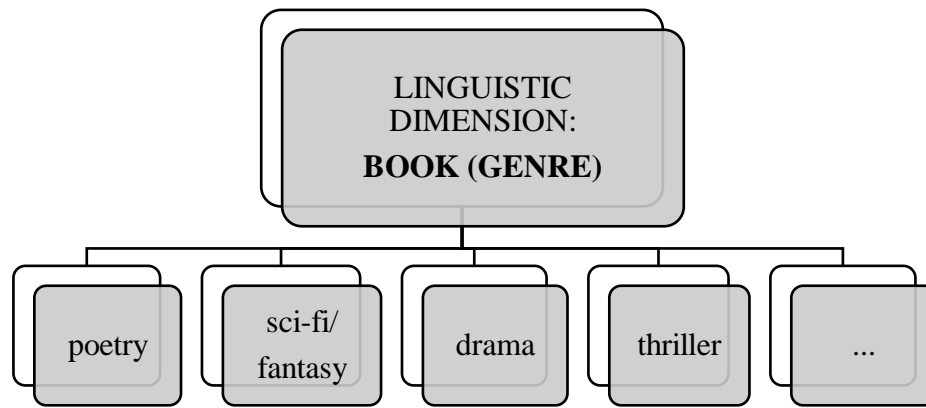


Figure 3. Compiled by author.

- Prominence: Also referred to as ‘the distribution of attention’. According to Talmy (2006), within the subdivision of prominence are the categories of level of synthesis and level of exemplarity, as well as the component of the perspectival mode category that involves global vs. local scope of attention. Gibbs. et al. describe four important image schema transformations: (1) path-focus –movement of an object– to end-point focus –resting point–, (2) multiplex to mass, (3) superimposition – changing the size of objects to fit within the other–, and, lastly, (4) following a trajectory. Langacker describes this last transformation with a trajector-landmark approach: ‘trajector’ refers to the primary focus, while ‘landmark’ refers to a secondary focus. He points out that the prepositions ‘above’ and ‘below’ are inherently similar as they profile the same relationship, but with a different focus.



The lamp is above the table.

The table is below the lamp.

Figure 4. Compiled by author. Source: Langacker R. W. (2008)

- Perspective/perspectivization: as the name indicates, this system refers to how someone chooses where to place their ‘mental eyes’ (Talmy, 2006). In essence, it is the position from which someone regards something, taking into consideration the prominence or distribution of attention. There are two main types: (1) steady-state long-range perspective (global scope of attention), and (2) moving close-up perspective (local scope of attention. For example (Talmy, 2006):
 - a) Global scope: “There *are* houses *at various points* in the valley.”
 - b) Local scope: “There *is a* house *every now and then through* the valley.”

3.4. Categorization

Lastly, categorization refers to what makes humans identify nouns as nouns, verbs as verbs, questions as questions, etc. Langacker stated that it is possible to define certain types of speech based on their meaning. This happens because the cognitive ability of categorization is a phenomenon that can be thought of as a ‘capacity of man’ (Geeraerts, 2006). This is what allows humans to have the cognitive capacity to construe events. For instance, a ‘concert’ is an abstract object/noun; he called this process conceptual reification. The words ‘resist’ and ‘resistance’ can serve as an example: while ‘resist’ directly reflects the process of the event, ‘resistance’ construes it as something that is abstract in nature, derived by conceptual reification. Conceptual contract is what makes words belong to different grammatical categories.

In Spanish, lexical roots establish a connection that create semantic families:

pan -	bread	
semantic family	panadería (n.)	bakery
	panadero (n.)	baker
	empanar (vb.)	to coat in breadcrumbs

Table 5. Compiled by author.

While, in Mandarin, this is manifested through the use of radicals:

人 (rén)	person	
亻	<i>character for person, transformed into a radical</i>	
uses	你 (nǐ)	you
	体 (tǐ)	body
	住 (zhù)	to live

Table 6. Compiled by author.

III. METHODOLOGY

This paper will provide an answer to the following questions:

- How does linguistic ambiguity manifest in Mandarin Chinese?
- How does linguistic ambiguity manifest in Spanish?
- How does each language resolve the different types of linguistic ambiguity?

The core objective of the analysis will be to provide an appropriate answer to the above stated questions; all of this is done in order to compare both. All languages in the world present linguistic ambiguity. However, Mandarin and Spanish are two that show it in different ways. This is due to their origins and how they developed in time.

In this manner, the paper aims to provide enough examples of all three types of linguistic ambiguity –lexical, semantic, and structural– of both languages to compare them. For this, the next section first offers a brief introduction to ambiguity. For the completion of this study, a variety of documents were consulted, accessed, and referenced. Secondly, the analysis was carried out mostly using the theoretical framework as a starting point.

Firstly, this paper offers an in-depth framework of essential concepts related to the topic; a variety of documents –journals, reports, websites, etc.– were consulted, accessed, and referenced. These documents were mostly accessed through Google Scholar, Research Gate, Academia, and JSTOR. Secondly, the analysis was carried out mostly by consulting a variety of Standard Mandarin Chinese and Spanish dictionaries, but the ones that helped to the completion of this paper were the *Collins Dictionary* for Mandarin and the *Diccionario de la Real Academia Española* (DRAE) for Spanish.

The analysis will now aim to answer the above stated questions.

IV. ANALYSIS

The analysis will begin by exploring the three different types of ambiguity that are present in both Mandarin and Spanish; this will be done by using cognitive grammar as a starting point. Linguistic ambiguity is a quality of language that makes spoken speech or written text open to interpretation; for the correct interpretation of a message, the receiver requires additional contextual information. This is a quality that presents a big difficulty, especially for those who are learning a language. Linguistic ambiguity can be classified into three main types: lexical, semantic, and structural.

1. Lexical Ambiguity

Lexical ambiguity occurs when a word or a phrase has multiple meanings. In Mandarin, lexical ambiguity is especially common in spoken speech. As previously mentioned, Mandarin has five tones, which serve to distinguish different meanings. However, Mandarin has multiple homophones, these are words that have the exact same tone and only differ in written language. The following table presents of examples.

记忆	≠	技艺
jìyì	=	jìyì
to remember	≠	skill
钟	≠	终
zhōng	=	zhōng
clock	≠	end, death

悲剧	≠	杯具
bēijù	=	bēijù
tragedy	≠	glassware
蝠	≠	福
fú	=	fú
bat	≠	good fortune, happiness
目的	≠	墓地
mùdì	=	mùdì
goal	≠	graveyard, burial ground
报复	≠	抱负
bàofù	=	bàofù
to retaliate	≠	ambition

Table 7. Compiled by author. Source: Collins Dictionary.

In addition, there are words that, although they may not have the same tone, they sound similar, which can lead to confusion. To give an example:

四	≠	死
sì	=	sǐ
four	≠	to die, dead

Table 8. Compiled by author. Source: Collins Dictionary.

It must be mentioned that, while there are words in Mandarin that are written exactly the same, but pronounced differently, there are some cases of these; they are called homographs. In this language, it is easier to understand the real context when words are in written form.

得	=	得
de	≠	děi
<i>aux. particle</i>	≠	need, must
了	=	了
le	≠	liǎo
<i>particle to indicate the past tense</i>	≠	<i>verb: finish, understand, etc.</i>

Table 9. Compiled by author. Source: Collins Dictionary.

Contrary to Chinese, since written Spanish uses the Latin script and it is alphabetical and phonemic, the concept of a written word being pronounced differently is illogical –this does not include different accents. However, the Spanish language does have a great quantity of homophones, these are some examples transcribed phonetically based on the International Phonetic Alphabet:

vaca	≠	Baca
/ba ka/	=	/ba ka/
cow	≠	roof rack

hola	≠	ola
/o la/	=	/o la/
hello	≠	wave
bello	≠	vello
/be lo/	=	/be lo/
beautiful	≠	hair
bienes	≠	vienes
/bje nes/	=	/bje nes/
property	≠	to come (<i>conjugated form</i>)
basta	≠	vasta
/bas ta/	=	/bas ta/
enough	≠	vast
honda	≠	onda
/on da/	=	/on da/
deep	≠	wave

Table 10. Compiled by author. Source: Diccionario de la Real Academia Española.

Some Spanish words are pronounced and written the same. However, they have different meanings; in spoken speech, context provides enough information as to what the real meaning may be. In written form, one of them will have an acute diacritic accent. To give some examples:

tu	=	tú
/tu/	=	/tu/
your	≠	you
si	=	sí
/si/	=	/si/
if	≠	yes
dé	=	de
/de/	=	/de/
to give (<i>conjugated form</i>)	≠	of, from, etc.

Table 11. Compiled by author. Source: Diccionario de la Real Academia Española.

It is through the cognitive ability of association that there is an established connection that creates these words and gives them meaning. In these cases, there is a need of a written phonological pole in order to know what the semantic pole truly is. In particular, this is especially the case of Mandarin.

2. Semantic Ambiguity

Semantic ambiguity, also known as polysemy, occurs when a word has more than one meaning. There are those who have proposed that polysemy is a pragmatic phenomenon caused by the use of a language in a variety of contexts (e.g., Falkum, 2011, 2015, Falkum & Vicente, 2015), while others have suggested that it is a reflection of how rich the semantic content of lexical categories can be (e.g., Asher, 2011; Evans, 2015; Pustejovsky, 1998). In cognitive grammar, this can be described as part of the cognitive ability of association, in which meanings are associated to words.

Semantic ambiguity is extremely common in Spanish as the language does not have many words compared with other languages. In fact, while it is hard to determine the exact amount,

the *Diccionario de la Real de la Academia Española* (DRAE) has over 93,000 words. In the case of Mandarin, it is estimated to have around 500,000 words. However, the largest Mandarin Chinese dictionary has only over 370,000 words (辞海, cíhǎi). For this reason, in Spanish, there is an explicit need to use words in different contexts with different meanings to communicate. For instance:

<i>bolso</i> (n.)	<ol style="list-style-type: none"> 1. bag. 2. pocket. 3. a small handbag made of leather, cloth or another material, provided with a closure and frequently with a handle, used in particular by women to carry money, documents, objects for personal use, etc. 4. pocket of air formed when the wind hits the sails, which are being maneuvered.
<i>móvil</i> (n.)	<ol style="list-style-type: none"> 1. cell phone. 2. that which moves something materially or morally. 3. articulated sculpture, whose parts can be movable. 4. number assigned to each cell phone.

Table 12. Compiled and translated by author. Source: Diccionario de la Real Academia Española.

This polysemy was automatized by the Spanish speakers, as there was an explicit need to communicate certain things that did not have words per se.

Although Mandarin does have a good amount of polysemy, it is not as common as it is in Spanish, the reason being that this language has more words that allow speakers to communicate with a higher degree of precision. However, there is a phenomenon that is very common in Mandarin: polysyllabic words have one or more characters that are ‘reused’ and are there to provide a category. For instance:

机 (jī)	machine	
uses	耳机 (ěrjī)	earphones
	飞机 (fēijī)	airplane
	手机 (shǒujī)	mobile phone

Table 13. Compiled by author. Source: Collins Dictionary.

Furthermore, Mandarin uses radicals on each character, which allows the reader to understand what a character may refer to in case they do not know how to read it. Radicals are something that can somewhat help erase this type of ambiguity to an extent. For example:

水 (shuǐ)	water	
氵	<i>character for water, transformed into a radical</i>	
uses	河 (hé)	river
	海 (hǎi)	sea
	汗 (hàn)	sweat

Table 14. Compiled by author. Source: Collins Dictionary.

In this case, all of these characters have to do with water, that is why they carry that radical.

Speakers of two or more languages may encounter an additional ambiguity. For instance, the word ‘fin’ in Spanish has a different meaning to the English word. This is called interlingual homographs or false friends. However, Mandarin and Spanish do not share any false friends as both languages have very different origins, and they use distinct writing styles.

In addition, both Mandarin and Spanish are languages that rely on the cognitive ability of categorization to give a function to their words. Moreover, in Spanish, words can be derived from others, this creates a semantic family, while in Mandarin, words are not derived per se. However, they can have the same character as other words, which creates a connection between them. This also happens with the characters radicals.

3. Structural Ambiguity

When it comes to structural ambiguity, Spanish is a language that particularly shines. In the section that talks about the Spanish language within the theoretical framework, there is a table that goes over the many possibilities that the language allows to create a sentence. Mandarin does have structural ambiguity, but only to a limited extent. Here are two examples in both Spanish and Mandarin:

Vi a la niña con gafas.

This example can be understood in two ways depending on how the attention is distributed:

- I saw the girl, who was wearing glasses.
- I was wearing glasses, so I saw the girl.

The meaning of this sentence changes depending on who is the trajectory and who is the landmark.

我看到那个戴眼镜的女孩。

wǒ kàn dào nàgè dài yǎnjìng de nǚhái.

This example is a direct translation of the Spanish sentence. However, because Mandarin requires a higher degree of precision, the verb 戴 (to wear) has to be present, otherwise, the

sentence does not make sense –without it, the meaning of this sentence would be something like “I saw the glasses’ daughter”, which is illogical.

It is through the cognitive ability of categorization that there is a possibility to understand the meaning of a sentence when there is structural ambiguity. This is the process that allows humans to recognize nouns as nouns, verbs as verbs, etc. In syntax, “con gafas” is a prepositional phrase (PP) and depending on what it is associated to, it will have a different meaning. For this reason, humans who speak languages with high structural ambiguity have automatized the need for context, otherwise, they will most likely assume that the prepositional phrase refers to the girl and not to the subject (I). As a hypothesis, this may be because of the word order and simply because the PP goes after “la niña”, it is understood that it is the girl the one who is wearing glasses; in other words, a matter of proximity.

In addition, Mandarin does require a certain amount of context to understand when an action is taking place, as the language does not have verbal tenses per se. However, Spanish has a total amount of 23 tenses that allows the speaker to segment time based on different situations, conditions, etc. For example:

Se había ido a casa.

他已经回家了。

tā yǐjīng huíjiā le.

These two sentences share the same meaning. However, in Spanish, the use of the verbal tense past perfect or ‘pluscuamperfecto’, transmits the idea of ‘already’, while in Mandarin, there is an explicit need of context: first, to imply ‘already’ (已经), and, second, to imply that it is a past action (了).

V. SUMMARY AND CONCLUSION

On the one hand, in Mandarin lexical ambiguity is especially common in spoken speech, seeing as it has five tones. This language also has multiple homophones, these are words that have the exact same tone and only differ in written language. For example, 钟 and 终 are both pronounced as ‘zhōng’, but the former means ‘clock’ while the latter means ‘end’ or ‘death’. As a side note, this is the reason why, in China, it is considered bad manners to gift someone a clock as it can imply that death is being wished upon the receiver. For this reason, in this language, it is easier to understand the real context when words are in written form. It must be mentioned that Mandarin has a number of words that share the same character, but are pronounced differently, which can make the language quite ambiguous in its written form without a proper amount of context. Mandarin has a good amount of polysemy, but it is not as common as it is in Spanish, the reason being that this language has a bigger number of words that allow speakers to communicate with a higher degree of precision. Because Mandarin requires a higher degree of precision, the language requires a certain amount of context to understand when an action is taking place, as the language does not have verbal tenses per se.

On the other hand, Spanish is alphabetical and highly phonemic. Therefore, the concept of a written word being pronounced differently is illogical. However, the Spanish language does have a great quantity of homophones, for example, both ‘vaca’ and ‘baca’ are pronounced as ‘/ba ka/’, yet the former refers to the animal cow, and the latter means roof rack. Moreover, semantic ambiguity is common in Spanish, seeing as the language does not have as many words as Mandarin, for instance. Therefore, the language has developed in such a way that words are used in different contexts with different meanings to communicate.

When it comes to structural ambiguity, Spanish is a language that particularly shows this type of ambiguity; from the order of words to the absence of some of these. People who speak languages with high structural ambiguity have automatized the need for context.

All of these findings have proven the hypotheses that were proposed in the methodology section. To sum them up, Mandarin’s tones add a certain degree of linguistic ambiguity. Even without having the same tones, words can be confused in spoken speech. Written speech usually helps erase ambiguity to an extent due to each character being highly specific. Spanish linguistic ambiguity relies more on polysemy and the diversity of language structure, which can give way to a myriad of different meanings.

In addition, this paper also offers an answer to all of the question that were posed in the methodology section. The following conclusions are presented:

- Mandarin's diversity of tones adds linguistic ambiguity in spoken speech, which means that ambiguity could potentially difficult understanding among native speakers and also the learning of pronunciation for non-native speakers.
- Spanish's linguistic ambiguity is more rooted in polysemy and the diverse ways a sentence can be constructed. This could potentially difficult understanding among speakers and learners.
- On the one hand, Mandarin characters are able to solve linguistic ambiguity to an extent, but this language's lack of verbal tenses can add linguistic ambiguity, especially for non-native speakers. On the other hand, Spanish diacritic accent helps solve linguistic ambiguity in written speech. In addition, Spanish's number of verbal tenses erases linguistic ambiguity to an extent on a structural level.

Moreover, the extension of this paper is limited by its nature, seeing as it is an undergraduate final degree project. For this reason, this specific topic of research could be extended in a number of ways mostly because there are not many cases that compare these two languages. Each kind of linguistic ambiguity could be expanded and substantiated in more depth to achieve a higher understanding of how linguistic ambiguity is manifested in both languages and how each of them attempt to solve it. Future research could also include the impact of the differences for the language learner and how comparative cognitive grammar might provide solutions.

VI. BIBLIOGRAPHY

- ASHER, N. (2011). *Lexical Meaning in Context: A Web of Words*. Cambridge: Cambridge University Press.
- BJÖRKSTÉN, J. (1994). *Learn to Write Chinese Characters*. Lund: Yale University Press..
- CHAO, Y. R. (1968). *A grammar of spoken Chinese*. Berkeley: University of California Press.
- COLLINS DICTIONARY (n.d.). *Collins Chinese Dictionaries*. Retrieved from Collins: <https://www.collinsdictionary.com/dictionary/english-chinese>
- EVANS, V. (2015). A Unified Account of Polysemy within LCCM Theory. *Lingua*, 152(6), 201-224.
- FALKUM, I. L. (2011). *The Semantics and Pragmatics of Polysemy: A Relevance-Theoretic Account*. London: University College London.
- FALKUM, I. L. (2015). The How and Why of Polysemy: A Pragmatic Account. *Lingua*, 157(1), 83-99.
- FALKUM, I. L., & VICENTE, A. (2015). Polysemy: Current Perspectives and Approaches. *Lingua*, 157(6), 1-16.
- GEERAERTS, D. (2006). Prototype Theory. In D. Geeraerts, *Cognitive Linguistics: Basic Readings* (pp. 141-166). Berlin: Mouton de Gruyter.
- GIBBS, R. W., & COLSTON, J. (2006). Image schema. In D. Geeraerts, *Cognitive Linguistics: Basic Readings* (pp. 239-268). Berlin: Mouton de Gruyter.
- INSTITUTO CERVANTES. (2022). *El español en el mundo: anuario del Instituto Cervantes*. Madrid: Instituto Cervantes.

- INTERNATIONAL COGNITIVE LINGUISTICS ASSOCIATION. (n.d.). *About Cognitive Linguistics: Historical Background*. Retrieved from International Cognitive Linguistics Association: <https://www.cognitivelinguistics.org/en/about-cognitive-linguistics>
- LANGACKER, R. W. (2008). *Cognitive Grammar: A Basic Introduction*. Oxford: Oxford University Press.
- LAPESA, R. (1981). *Historia de la lengua española*. Madrid: Editorial Gredos, S.A. .
- MERRIAM-WEBSTER DICTIONARY. (n.d.). *Meaning of the word "mandarin"*. Retrieved from Merriam-Webster Dictionary: <https://www.merriam-webster.com/dictionary/mandarin>
- MUÑOZ MACHADO, S. (2019, March 26). *Cómo América Latina asumió la lengua castellana o española como propia*. Retrieved from WMagazín: <https://wmagazin.com/relatos/como-america-latina-asumio-la-lengua-castellana-o-espanola-como-propia/#hablamos-la-misma-lengua>
- NORMAN, J. (1988). *Chinese. Cambridge language surveys*. Cambridge: Cambridge University Press.
- PENNY, R. J. (2004). *A History of the Spanish Language*. Cambridge: Cambridge University Press.
- PUSTEJOVSKY, J. (1998). The Semantics of Lexical Underspecification. *Folia Linguist*, 32(3/4), 323-347.
- REAL ACADEMIA ESPAÑOLA; ASALE. (2010). *Ortografía de la lengua española*. Espasa.
- REAL ACADEMIA ESPAÑOLA (n.d.). *Diccionario de la lengua española*. Retrieved from Real Academia Española: <https://dle.rae.es>

- SAGART, L. (2006). The origin of Chinese tone. In *Proceedings of the Symposium Cross-Linguistic Studies of Tonal Phenomena, Tonogenesis, Typology, and Related Topics*. Paris: HAL.
- SALA, M. (n.d.). *Vulgar Latin*. Retrieved from Britannica: <https://www.britannica.com/topic/Vulgar-Latin>
- STATISTA RESEARCH DEPARTMENT. (2023, March 31). *The most spoken languages worldwide 2022*. Retrieved from Statista: <https://www.statista.com/statistics/266808/the-most-spoken-languages-worldwide/>
- TALMY, L. (2006). Grammatical Construal. In D. Geeraerts, *Cognitive Linguistics: Basic Readings*.
- THE EDITORS OF ENCYCLOPAEDIA BRITANNICA. (n.d.). *Jin dynasty: China-Mongolia*. Retrieved from Britannica: <https://www.britannica.com/topic/Jin-dynasty-China-Mongolia-1115-1234>
- THE EDITORS OF ENCYCLOPAEDIA BRITANNICA. (2022, November 15). *Song dynasty*. Retrieved from Britannica: <https://www.britannica.com/topic/Song-dynasty>
- THE EDITORS OF ENCYCLOPAEDIA BRITANNICA. (2022, December). *Tang dynasty*. Retrieved from Britannica: <https://www.britannica.com/topic/Tang-dynasty>
- THE EDITORS OF ENCYCLOPAEDIA BRITANNICA. (2022, November 28). *Zhou dynasty*. Retrieved from Britannica: <https://www.britannica.com/topic/Zhou-dynasty>
- UNITED NATIONS. (n.d.). *What are the official languages of the United Nations?* Retrieved from Ask UN: <https://ask.un.org/faq/14463>
- WENG, J. (2018, August). What Is Mandarin? The Social Project of Language Standardization in Early Republican China. *The Journal of Asian Studies*, 77(3), 611-633.

VII. ANNEX

1. Mandarin

Mandarin belongs to the Sino-Tibetan language family, which directly descends from Proto-Sino-Tibetan (reconstructed proto-language). Eventually, Mandarin began to spread out and evolved into a series of languages. In this manner, Old Chinese came to be. The earliest examples of Chinese date back to 1250 BCE and they were written by using the Oracle Bone Script.

A more recognizable form of Old Chinese, this one is also known as Classical or Literary Chinese (古文 - guwen), developed during the Zhou dynasty (1046 BCE-256 BCE), known for the establishment of certain political and cultural characteristics that would be identified with China for the next two millennia. Although it has been a subject of controversy, the overall consensus is that Chinese had not yet developed tones (Sagart, 2006). Classical Chinese has been used as the formal written language until the early 20th century. However, spoken Chinese did evolve into what is known as Middle Chinese (中古汉语 - zhongguyanyu); this is the ancestor of almost all Modern Chinese varieties.

Middle Chinese was not a single unified language, but a number of mutually intelligible dialects. During this time, alongside Classical Chinese, Written Vernacular Chinese (白语文) also developed to reflect spoken Chinese in its different varieties. This primarily took place during the Tang (618-907 CE) and Song (960-1279) dynasties. Also during this period, tones are thought to have begun to appear in different ways (Sagart, 2006).

Old Mandarin (古官话 - guguanhua) was a language, which can be recognized as a form of Mandarin, that was spoken during the Jin (1115-1234) and Yuan (1271-1368) dynasties. The Yuan dynasty is particularly important to the development of the language as it was a long period of Mongol control of China. The war leading up to the Mongol conquest caused a large-scale migration of Old Mandarin speakers to the South and, subsequently, spread early Mandarin dialects to more areas, as well as had a heavy influence on other emerging Chinese languages (Norman, 1988).

During the Ming (1368-1644) and Qing (1644-1911/1912) dynasties, an official administrative language was created primarily based on the Old Mandarin dialect that was

spoken in the southern capital, Nanjing. This was the first instance at which the language was referred to as ‘Mandarin’; now considered to be Middle Mandarin (官话 – guanhua). The Qing dynasty was the last of the imperial dynasties of China, during this time, the population of the empire grew exponentially and, therefore, the number of dialects did too. In the late 19th century, the dialect spoken in Beijing became particularly important; it replaced the variety of the Nanjing dialect, and became the new high variety (Norman, 1988).

At the beginning of the 20th century, a new official written language was introduced in order to replace Classical Chinese. However, it was later decided that the official language should be based on the Beijing dialect due to its importance. In this manner, Putonghua –what is now known as Standard Mandarin Chinese or, simply, Mandarin– was born.

2. Spanish

The Spanish language belongs to the Indo-European family of languages; the first ancestor can be dated back to approximately 5,000 years ago around the area of the Black Sea. Ultimately, those who spoke Indo-European started to spread in various directions and, gradually, varieties of it began to appear. In this case, there is a particular branch that is can be traced as the origins of Spanish, and that is the Italic branch.

Latin became the most prominent member of the Italic branch by means of political and cultural pressure from Ancient Rome. As Rome was spreading out and conquering more territories, Latin also did due to Roman soldiers settling in those territories. However, these soldiers did not speak Classical Latin (i.e. literary Latin), but what is known as Vulgar Latin. Vulgar Latin was less standardized in its grammar and vocabulary and is, essentially, “a spoken form of non-Classical Latin from which originated the Romance group of languages” (Sala, n.d., para. 1).

During and after the Roman conquest of the Iberian Peninsula, as previously stated, soldiers settled there. Although Latin was not enforced, the local populations took the decision to learn it out of convenience and, in some cases, prestige; this process was called Latinization. Generally, it is believed that this phenomenon began due to Romanization with started at the beginning of the Second Punic War, in 218 BC, and it “was relatively slow, progressing in a westerly and northwesterly direction over the next two centuries and culminating in the

conquest of the northern coastal area [of present-day Spain]” (Penny, 2004, p. 8). Furthermore, the end of the Roman state brought with it a reason of sorts to adhere and perpetuate the incorrect features of Vulgar Latin. It should be mentioned that other Roman territories did speak Vulgar Latin too, however, all of them evolved and developed in different ways.

In time, different dialects descending from Vulgar Latin did also emerge within the Iberian Peninsula, more specifically what is now known as Spain, and slowly began to become distinct languages; the most important ones are: Aragonese, Castilian, Catalan, Galician-Portuguese, and Portuguese. In particular, Castilian arose in the north central region, in the province of Castille.

Alfonso X the Learned, king of Castille and Leon (1252-1284) pushed the standardization of Castilian as a distinct language for its official use, based on the dialect spoken in Toledo. He was the one who established a spelling system that successfully specified vernacular pronunciation making it fairly phonemic.

Due to the presence of Arabic in the peninsula, and especially during and after the last centuries of the Reconquest (722-1492), Castilian was heavily influenced by Mozarabic dialects; it is said that Castilian absorbed around 4,000 words (Lapesa, 1981). After 1492, Spain sent explorers –like Christopher Columbus– abroad to help build the Spanish Empire. Castilian spread along with the empire to the Americas, the Spanish East Indies, and some parts of Africa. American Spanish gradually came to be, as there was a pressing need to understand the natives and with time, Castilian Spanish also absorbed some words from native south American languages (Muñoz Machado, 2019). Furthermore, although to a lesser degree, Tagalog did also have an influence on Castilian; however, the influence of present-day Tagalog is much greater.

“It is in the 16th century that the term [Spanish] is applied to the language of culture of Spain, and therefore becomes equivalent in the sense to [Castilian]” (Penny, 2004, p. 31). In this manner, Spanish was born and became a world language, while Castilian was reduced to the specific name of the variation of Spanish spoken in peninsular Spain.