

AN APPROACH TO THE LEXICAL AMBIGUITY CAUSED BY FALSE COGNATES IN SPANISH L2. A CORPUS-BASED EXPLORATORY STUDY

María Sampedro Mella 

Abstract. This exploratory study aims to offer new insights into the lexical ambiguity in the interlanguage arising from the use of false cognates. The study is based on written texts from two learners' corpora. Utterances including false cognates were retrieved from both corpora and presented to two groups of native speakers who assessed their correctness and stated whether or not they were ambiguous. While the first group was informed beforehand that the sentences had been produced by non-native speakers (as well as the meaning of the false cognates in their L1), the second group had no information regarding the speakers who wrote them. The goal of this analysis is to shed light on the different perceptions that can arise from native vs. non-native speakers' productions, focusing on lexical ambiguity.

1. Introduction

False cognates have been analyzed extensively in studies of lexicography, language teaching and learning, translation, and contrastive linguistics. Indeed, L2 studies highlight the difficulties that learners encounter when using these kinds of words (e.g. Roca Varela 2010). However, to the best of my knowledge no study has investigated in detail how false cognates can create ambiguity among hearers. This may occur when a false cognate is apparently correctly used in an utterance, but can be interpreted in two different ways: with the meaning in the L1, and in the L2 of the speaker. This could be the case, for example, with the use of the false cognate *actualmente* [currently] in Spanish by an English speaker in a context where the word can mean both *actually* and *currently*, as in *El mayor problema que tenemos actualmente/en realidad es la falta de dinero* [The biggest problem we currently/actually have is the lack of money]. In such case, a Spanish native hearer who has knowledge of English may feel that the sentence is ambiguous, i.e., it can be understood in more than one way. This would be due to the fact that it was said by an English speaker who may not use the word correctly in the target language.

In order to carry out the present study, some utterances with false cognates were retrieved from the *Corpus de Aprendices de Español* [Spanish Learner Corpus] (CAES) of the Instituto Cervantes and the Universidade de Santiago de Compostela (Spain), and the *Corpus Escrito*

2 María Sampedro Mella

del Español L2 [L2-Spanish Written Corpus] (CEDEL2) of the Universidad de Granada (Spain). A questionnaire that contained eight items and two distractors was created based on these utterances, and presented to two target groups of native speakers, who assessed their correctness and indicated when they were ambiguous. The control group did not have any information on the speakers who produced the sentences. However, the critical group was informed that the sentences were produced by non-native speakers and contained some false cognates, whose meanings in the speakers' L1 were also provided. The main focus of this study is to determine if there is a different perception in the interpretation of an utterance when it is produced by a native or a non-native speaker, focusing on the lexical ambiguity.

This paper is organized in 5 sections: section 2 contains a theoretical background; section 3 presents the objectives and the research questions of this study; section 4 includes the methods, i.e., the description of the corpus data, the participants, and the design of the task; section 5 presents the results and the discussion; lastly, section 6 contains the conclusion.

2. Theoretical background

2.1. *Native language vs. interlanguage: the error*

The interlanguage is defined by Selinker (1972:214) as “a separate linguistic system based on the observable output which results from a learner’s attempted production of a TL [target language] norm”. This linguistic system has a different structure from the learner’s native and target languages, and contains linguistic errors caused by L1 transfer (*interlingual errors*) and by difficulties in the L2 learning process (*intra-lingual errors*) (Corder 1967). In a broader category of errors, Corder (1967) also makes a distinction between *mistakes*, i.e., those that the learner is able to self-correct, and *errors*, which reveal the learner’s lack of knowledge.

Errors have been regarded in many different ways in the learning process throughout the history of foreign language teaching. In traditional methods, errors were judged negatively, and they were deemed to affect the learning process of the target language. However, the interlanguage theory (Selinker 1972) and the error analysis approach (Corder 1967) presented a different view of the error, which was seen as a device used by learners and teachers when dealing with the L2: by the former when learning the L2, by testing out their hypothesis about the target language, and by the latter as evidence of the aspects of the L2 that are unclear to students. Today, the communicative language teaching approach considers that “errors are tolerated and are seen as a natural outcome of the development of communicative skills” (Larsen-Freeman 2011:159).

In the literature on Spanish as a foreign language (SFL) teaching, the error has been the subject of numerous classifications in various categories since the 1990s (see Penadés Martínez 2003 for further information). Furthermore, studies to date have adopted various perspectives, and have undertaken various procedures for analysis (Alba Quiñones 2009, Valverde Mateos 2012), such as contrastive analysis (e.g. Sampedro Mella 2021a, 2021b) and error analysis (e.g. Navarro Gala 1999, Belda Torrijos 2015, Sampedro Mella & Sánchez Gutiérrez 2017). Although the SFL literature has focused on grammatical and lexical errors, in recent years many authors have placed more importance on the discourse and the interactional level. In this context, numerous studies have focused on an analysis of the “pragmatic transfer” (Thomas 1983:91) displayed by L2 learners and the pragmatic failure.

In addition to linguistic and discourse errors, the interlanguage can likewise be subject to communication failures, due to non-native speakers not being able to communicate and being inaccurately understood. Indeed, even when communication is prioritized over linguistic structures, communication may occasionally stop at some point. In these cases, it is necessary for the speakers to clarify their messages or for the audience to reconstruct them: “Communication involves two speakers transferring meanings to each other using these preexisting sentences. In cases where there is a communication problem, a procedure of analytical reconstruction intervenes to recover the learner’s meaning obscured by his choice of the wrong sentence” (Hamilton 2001:81). For instance, examples (1) and (2) extracted from the CAES corpus illustrate a contradiction between what the learner says (or writes) and the logical thinking of the interlocutor which may affect the communication process¹:

- (1) *un ladrón está en qprieto y él decidí hacer algo para mantener la vida. Entró en un piso que está un barrio central y quiso llevarse algo. entró en la habición desde una ventana y miró los alrededores. la habitación está bien decorada, parece es perteneciente a un rico. ¡Genial! el ladrón pensó: ahora todos son mios. [...] Después de terminar todo, se le ocurrió una buena idea que quizás puede echar una siesta en una habitación tan grande, entonces se echó debajo de una cama y no tartó mucho en dormirse.* [L1-Chinese, B1; CAES corpus]

‘A thief is in trouble and he decided to do something to save his life. He came into an apartment which is in a central neighborhood and wanted to take something. He came into the room through the window and looked around. The bedroom was well decorated, seems to belong to a rich person. Great! The thief thought: now everything

¹ In this study, the examples in Spanish are shown as they appear in the corpus (with grammatical and orthographical errors). All translations are mine.

4 María Sampedro Mella

is mine (...) After finishing everything, he had the great idea of perhaps taking a nap in a big bedroom, so **he threw himself under the bed and it did not take long for him to fall asleep**'.

- (2) *Como la economía comenzó a bajar desde hace 10 años, la empresa decidió de disminuir el numero de trabajador. Al principio mi seccion estuvo muy bien [...] Pero la venta de los productos fui bajando poco a poco desde hace 3 años a pesar de que la economía de Japón recuperaba gradualmente y por fin yo perdí el trabajo hace medio año* [L1-Japanese, B1; CAES corpus]

'As the economy started going down ten years ago, the company decided to reduce the number of employees. In the beginning my sector was doing very well (...). But the sales of the products were little by little going down for three years although the economy of Japan was gradually recovering **and at last I lost my job half a year ago**'.

There is a contradiction between the reader's expectations and what the learner wrote in both examples. In example (1), it seems strange for a person to take a nap under the bed, so the choice of the adverb of place (*debajo* [under] instead of *encima* [on]) seems to be wrong. However, the fact that the person concerned is a thief who may want to hide from the owner of the apartment makes this situation more plausible. As we do not have any more information about this example, we could not guarantee whether this is an error or not.

In example (2), however, it is more likely to be an error in the choice of discourse marker: the learner uses a connector which expresses relief to state that he has lost his job. Although this could be a reason for feeling better (e.g. if the working conditions were not entirely favorable), the content of the text suggests that it is an error. Indeed, Ramos Sañudo (2021), based on a learner corpus-based analysis, observes that most Spanish L2 students do not notice the meaning of this discourse marker which is used as a closing discourse marker to indicate the final point of the text: "ha quedado patente la dificultad que existe con respecto a *por fin* y *al fin*, dos marcadores de cierre con un claro valor modal, que sin embargo no son percibidos como tales por los aprendices" [it is clear that there is a difficulty regarding *por fin* [at last] and *al fin* [after all], two closing markers with a clear modal value that are nevertheless not perceived as such by [L2 Spanish] learners] (Ramos Sañudo 2021:277).

It is interesting to note that examples (1) and (2) make sense, despite their grammatical and orthographical errors. If they were produced by a native speaker, they would leave no room for doubt and ambiguity. However, the fact that they were produced by a non-native speaker leaves room for doubt and a double interpretation: the hearers expect the speaker not to have full competence in the L2, and they are therefore more likely to commit errors when producing oral or written texts.

Furthermore, when the content of the message comes into conflict with the hearers' expectations, they will probably consider the text to be wrong or to contain an ambiguity, in that it could have more than one interpretation. The "ambiguity error" that can have a significant effect on the correct interpretation of a text is thus included in some classifications of learners' errors in the Spanish literature as a type of communicative error (cf. Vázquez 1991). However, in contrast to the syntactic ambiguity error due to the order of the words in the sentence (e.g. Fernández Solera 2002, Belda Torrijos 2015), there is no extensive bibliography about the lexical ambiguity error in SFL, or even in the literature on English as a foreign language².

2.2. Lexical ambiguity and interpretation

Ambiguity in languages refers to a linguistic phenomenon in which a phrase or a sentence has two or more definite meanings or readings. Phonetic, syntactic and lexical ambiguity are constantly present in languages (Ullmann 1976). In particular, lexical ambiguity is caused by the homonymy and the polysemy of the languages, i.e. the relationship between words with identical forms but different meanings, and the coexistence of many possible meanings for a word or phrase respectively (see Horno & López 2020 for further information and a didactic proposal).

In addition to the lexical ambiguity due to the homonymous and polysemic words, a third type must be considered in the interlanguage: the ambiguity caused by *false cognates*. *Cognates* are words that share a similar meaning, spelling, and pronunciation in two or more different languages. For example, the words *person* and *academic* share the same origin and meaning, but do not have exactly the same spelling in English, Portuguese (*persoa*, *académico*) and Spanish (*persona*, *académico*). In addition to the same origin and meaning, cognates sometimes also have the same spelling; this is the case of words like *particular* and *individual* in the three mentioned languages (for further information, see Stamenov 2009).

As languages contain many cognates, learners tend to transfer similar words from L1 into L2, or from an L2 into a third foreign language. This is an automatic and subconscious use of past learner behaviors in the attempt to produce new responses. Indeed, Wang (2021:253) highlights that "[i]n the process of learning a second

² There is extensive literature concerning the "ambiguous error" which is not related to the type of ambiguity described in this paper. In the English literature, "[a]mbiguous errors are those that could be classified as either developmental or interlingual errors because they reflect errors that an L1 learner commonly makes while at the same time reflect the structure of the L2 learner's mother tongue. An example of this would be *He no have time* for a Spanish L1 speaker learning English" (Shaffer 2005:3).

6 María Sampedro Mella

language, college students just try to find the equivalents in the native language. Thus, ambiguities appear in second language learning with the first language interference”. This interference may become an obstacle when instead of transferring a cognate from one language to another, L2 students choose a *false cognate*. According to Stamenov (2009:219–220),

False cognates refer to pairs of words in the same or different languages that are similar in form and meaning but have different roots, i.e., do not share a common origin. They are juxtaposed, on the one hand, to cognates and, on the other, to false friends (or *faux amis*) that are pairs of words in two languages that look and/or sound similar, but differ in meaning.

In a broader perspective adopted here, false cognates are false friends, i.e. words that are identical or similar in form in two or more languages (even if they do not share the same origin or etymons), but which differ in meaning. Following a semantic classification, they can be divided into “total” and “partial” false friends. “Total false friends imply a conspicuous semantic difference between the L2 and the L1: English and Spanish in this case (e.g. English *vase* vs. Spanish *vaso* [glass], English *avocado* vs. Spanish *abogado* [lawyer], English *robe* vs. Spanish *robo* [theft])” (Roca Varela 2010:718). Partial false friends present a semantic overlap, since two similar words have one shared meaning and one different meaning, due to the polysemy. For instance, the Spanish noun *circulación* means ‘blood circulation’, like the English word *circulation*, but it also means ‘people or traffic movement’ in Spanish.

In this context, the decoding of a message and the ability to understand and interpret a message may be influenced if the audience knows that it was produced by a non-native speaker, especially if the message in question seems to be unusual or unexpected, as in examples (1–2). False cognates may also lead to doubt among the hearers, and lead to them understanding a message in a completely different way from what the encoder was trying to convey, e.g. (3):

(3) *voy a hablar te de las vacaciones que he hecho este verano a Madrid y voy comenzar por decir que la ciudad es espantosa me ha gustado muchísimo visitar el museo del prado la casa real y lo Retiro [...] pero lo que más he gustado fue la noche de Madrid madre mia que guay, he hablado con un montón de personas interesantes [...] fueran unas vacaciones que me ha gustado mucho y quiero volver a Madrid [L1-Portuguese, A2; CAES corpus]*

‘I’m going to talk to you about the holidays I had this summer in Madrid and I’m going to start by saying that **the city is horrible**. I liked visiting the Museo del Prado, the Casa<sic> Real and the Retiro [park] a lot (...) but what I liked the most was the nightlife of Madrid. Oh my

God! So cool! I talked with a lot of interesting people (. . .) they were great holidays that I liked a lot and I want to go back to Madrid’.

This is clearly an error due to a L1 negative transfer: the word *espantoso* has exactly the same spelling in Spanish and Portuguese, but means ‘terrific’ in Portuguese (the L1 of the learner) and ‘horrible’ in Spanish (the target language). In this example there is no room for ambiguity, since there is a broader context which clarifies that the use of this adjective is undoubtedly an error. However, if the example only contains the first sentence (*la ciudad es espantosa*), there would be at least two possible scenarios: 1) the hearer would understand completely the opposite meaning, i.e. Madrid is horrible; 2) the hearer would notice some ambiguity, if they knew that the learner’s L1 was Portuguese and they were a Portuguese language user (e.g. a Spanish teacher in a Portuguese-speaking country). The fact that the speaker of the message is non-native and the hearer’s previous knowledge of their L1 may influence the hearer, in that they may find ambiguity in contexts where it would not exist if the text were produced by a native speaker. It is precisely these cases of “hearer’s ambiguity” due to false cognates that are at the heart of this paper.

3. Objectives and research questions

This study examines different ways in which native speakers interpret L2 Spanish utterances, regardless of whether they know that they were produced by non-native speakers. Two factors are considered: on the one hand, the possible ambiguity due to the use of a false cognate; on the other, the correctness of the utterances.

The participants were asked to rate ten sentences, of which eight contained a false cognate and two were distractors with a polysemic and a homonymous noun. One group of participants (the critical group) knew that the utterances were produced by L2-Spanish learners and contained some possible false cognates while the second group (the control group) did not have this information. The aim of this design is to respond to the following research questions:

- (1) Are Spanish speakers more inclined to find lexical ambiguity if they know that sentences have been produced by non-native speakers?
- (2) Do Spanish speakers rate sentences differently when they know that they have been produced by non-native speakers?

The main hypothesis is that hearers are more critical and more inclined to detect ambiguity when they know that speakers are non-native. It is therefore important to investigate whether participants have different

perceptions towards the speakers who made the utterances depending on whether they have this information or not.

4. Methods

4.1. *Corpus data: search and results*

This study is based on the *Corpus de Aprendizajes de Español (CAES)* (Spanish Learner Corpus) and the *Corpus Escrito del Español-L2 (CEDEL2)* (L2-Spanish Written Corpus). Both corpora contain written texts produced by learners of Spanish as a foreign language at various levels, from A1 to C1 (CAES) and from A1 to C1+ (CEDEL2), according to the *Common European Framework of Reference for Languages* (Council of Europe 2001). In their current forms, the CAES and CEDEL2 corpora contain written productions from learners of ten common native languages: Arabic, Mandarin Chinese, Japanese, French, Portuguese, Italian, Greek, English, German, and Russian. The CAES also includes data from L1-Polish learners and CEDEL2 from L1-Dutch learners. The CAES corpus currently contains over 970,000 words, and the CEDEL2 over 1,100,000 words.

The search tool for both corpora allows retrieval of statistical information and textual examples of elements, lemmas, word classes and grammatical categories, with filters on the parameters that make up the corpus (basically the learner's L1 and level of proficiency in Spanish, but also age, sex, country of origin, etc.). It is also possible to distinguish between lower and higher case words, and accented and non-accented words. However, neither CAES nor CEDEL2 include an error tag system which would facilitate retrieval of the different types of errors made by learners (see Rojo & Palacios 2016, for further information about the CAES corpus and Lozano 2022 for an overview of the CEDEL2 corpus).

When carrying out this study, the first step was to search in both corpora for a set of different words that are false cognates in Spanish and in another language (e.g. *success* in English vs. *suceso* [incident] in Spanish). After carefully reading all the output (i.e. the results), it was observed that most of the cases were clearly correct or incorrect utterances. For example: *Este mismo día, otro suceso llegó a suscitar de nuevo emociones trágicas* (L1-Greek) [The same day, another incident again provoked tragic emotions] vs. *Y intenta *procura por la madre, pero sin *suceso* (L1-Portuguese) [and [he] tries to *try [look for] the mother, but without *incident [success]]. Only a few instances could thus be considered ambiguous. There are a total of two English-Spanish and two Portuguese-Spanish false cognates; two of them are total false friends, while the other two are partial false friends (see Table 1):

Table 1. False cognates retrieved from the corpora.

English or Portuguese	Spanish	Total/partial	Meaning
Support (EN)	<i>Soportar</i>	Total	'Bear'
Attend (EN)	<i>Atender</i>	Partial	'Watch over'
Todavía (PT)	<i>Todavía</i>	Total	'However' (PT) 'still/yet' (SP)
Acordar(se) (PT)	<i>Acordar(se)</i>	Partial	'Wake up' (PT), 'remember' (PT, SP)

A total of eight ambiguous utterances in seven texts which contain these false cognates were retrieved from the corpora, and used to design the experimental task (3.2). It is important to note that these false cognates are also polysemic words in Spanish. For example, among other meanings, the Spanish verb *atender* means 'to pay attention', 'to help' and 'to watch over'; *acordar(se)* means 'to agree to' or 'to set a date' as well as 'to remember', etc. Unfortunately, ambiguous utterances containing false cognates that were neither polysemic nor homonymic words in Spanish were not found in any corpus at all. However, in spite of these additional meanings, the utterances only have one interpretation, in addition to the possible ambiguity derived from the use of the false cognate.

4.2. Participants and task

Forty-six native speakers of Spanish participated in this study. Thirty-six were females and ten were males, and their mean age was 21 years old, with a minimum of 20 and a maximum of 28. All of them were enrolled on linguistic courses at the Universidade de Santiago de Compostela in Spain. They were divided into two different independent groups, and the researcher carried out all the data collection, in order to ensure that the data were all collected in exactly the same way. The participants completed a task and a demographic questionnaire during normal class time and took on average 5–10 minutes to complete it. They were not informed of the aim of the research until the survey had been completed by all the students.

The demographic questionnaire included a series of questions that aimed to gather information about the participants' age, gender, academic background and nationality. The task contained ten utterances that were presented to them. As mentioned above, eight were real sentences extracted from CAES and CEDEL2 (see 4.1.), whereas the other two were distractors that were created *ad hoc* with several words containing lexical ambiguity: *solo*, which means 'only' and 'alone', and *banco*, which can be both 'bench' and 'bank'. The sentences were *Estaré trabajando solo por la mañana* [I will be working only/alone in the morning] and *Luis y Marta se citaron en el banco de la esquina* [Luis and Marta met at the bench/bank on the corner].

10 María Sampedro Mella

In order to avoid biases or influence on the participants, the original sentences retrieved from the corpora were corrected and reworded when they had grammatical or orthographical errors: e.g. **suportaba* was changed for *soportaba* [bear], *una conferencia *internacionales* was changed for *una conferencia internacional* [an international conference], etc.

The task was divided into two parts: the goal of the first part was to assess the correctness of each sentence based on an assessment scale from 0 (completely wrong sentence) to 3 (completely correct sentence); the second part consisted of indicating whether or not the sentence was ambiguous. If it was ambiguous, the participants were asked to circle all the words that they thought that had more than one meaning in those utterances. In some cases, there was more than one sentence, as an additional one was provided as a context. As indicated in the task instructions, they had to focus on the target sentence (highlighted in bold).

The critical group (25 participants) was informed that these sentences were written by non-native speakers. They were also provided with a list of words and their meaning in the speakers' L1 (e.g. *todavía-however* in Portuguese). The control group (21 participants) completed the task with no information about the speakers who had written the sentences.

5. Analysis and discussion

5.1. Corpus data

Examples (4) to (7) contain the original texts in Spanish produced by different L1 learners, which were downloaded from CAES and CEDEL2 corpora. The literal translations into English and an alternative interpretation due to the meaning of the false cognates in English (4–5) and Portuguese (6–7) are also provided. The false cognate is underlined both in Spanish and in the translation in each sentence:

(4) *Soportar*

- a. *los fumadores tienen que respetar los demás y fumar en un sitio lejano de la gente es decir solos. **No somos obligados de suportar la decisión del otro.*** [L1-Arabic, B2; CAES corpus]
'smokers must respect others and smoke in a distant place from people, that's being by themselves. **We are not obliged to bear/support the other's decision.'**
- b. *Biografía de mi abuela Shajira*
Nací en Palestina (no recuerdo el año, hace mucho tiempo!) [...]
*Ella cocinaba muchas comidas ricas y quería compartir con toda la familia. Ella estaba muy fuerte pero sin agresión y siempre escuchaba y buscaba para soluciones. **Ella suportaba sus hijos y hijas siempre y nunca gritaba.*** [L1-Arabic, A2; CAES corpus].

‘Biography of my grandmother Shajira.
She was born in Palestine (I don’t remember the year, it was a long time ago!) (. . .) She cooked many tasty dishes and wanted to share everything with the family. She was very strong but not aggressive? and always listened and looked for solutions. **She always bore/supported their sons and daughters** and never shouted.’

(5) *Atender*

a. *Una semana no dio abasto para descubrir todos los misterios de Kashmir. Sin embargo, fuimos que regresar ya que **teníamos que atender la celebración de matrimonio de mi única prima**. [L1-English, C1; CEDEL2 corpus]*
‘A week was not enough to discover all the mysteries of Kashmir. However, we had to come back, since **we had to watch over/attend the marriage celebration of my only cousin.**’

b. *Este verano, fui a Japón con mis amigos para atender en una conferencia internacionales. Era una conferencia de noreste Asia, por eso, hay los estudiantes de Japón, Taiwán, Corea y Mongolia [L1-Chinese, B1; CAES corpus]*
‘This summer **I went to Japan with my friends to watch over/attend an international conference.** It was a conference in northeast Asia, that’s why there are students from Japan, Taiwan, Korea and Mongolia’

(6) *Acordar*

a. *¡Hola mi amigo!
¿Que se pasa? Como ayer acordé tarde, no te escribí sobre mi familia [L1-Portuguese, A1; CAES corpus].*
‘Hello my friend!’ ‘What’s up? As yesterday **I remembered/woke up late**, I didn’t write you about my family’

b. *Tenía que estar en aeropuerto a *leçças* 07:00, pero esta fue la hora que acordé [. . .]*³
‘I had to be at the airport at 07:00, **but that was the time I remembered it/woke up.** (. . .)’

(7) *Todavía*

a. [. . .] *Entonces tuve que cambiar de vuelo e **pasé el día intentando hacerlo por teléfono, todavía no lo logré.** [L1-Portuguese, A2; CAES corpus]*
‘(. . .) So I had to change the flight and I spent the day trying to do it by phone, **I haven’t yet got it/however I haven’t got it.**’

b. *Una de las personas que más admiro en mi familia, además de mis padres, es mi hermana. Ella se llama Eduarda, nació en Famalicão,*

³ The example continues in (7) a.

12 María Sampedro Mella

el lugar que vivimos actualmente, y nuestra familia es muy pequeña. Tenemos cuatro tíos y cinco primos, todavía, ya no tenemos ni abuelos ni abuelas. [L1-Portuguese, A2; CAES corpus]

‘One of the people from my family that I admire the most, aside from my parents, is my sister. Her name is Eduarda, she was born in Famalicão, the place where we currently live, and our family is very small. **We still have four uncles and five cousins, / We have four uncles and five cousins, however,** we don’t have neither grandfathers nor grandmothers anymore.’

Interestingly, while Portuguese-Spanish false cognates are only used by L1-Portuguese learners (6–7), English-Spanish false cognates are also used by other L1 speakers, including L1-Chinese (5b) and L1-Arabic (4) learners. This is because many students learn Spanish as a third or fourth language through another foreign language, typically English or French (for further information, see for example the learners’ notes in CAES corpus about the languages that they speak and learn).

The sentences in (4–7) a. are coherent and clear in Spanish, but they could be equally possible with the meaning of the false cognates in the speakers’ L1 (or possibly in a different L2), in that both meanings are quite similar. This is true of (7) a. for example, in which the speaker was ultimately unable to change their flight: *I haven’t yet got it* [Spanish meaning]; *however, I haven’t got it* [L1 meaning]. In (5) a. the speaker states that *we had to come back, since we had to watch over* [Spanish meaning] / *attend* [L1 meaning] *the marriage celebration of my only cousin.* In this context, both sentences imply being at the wedding.

Sentences may sometimes have a completely different meaning because of the use of false cognates, as in (3). In (6) a., regarding the partial false friend *acordar* (‘to remember’ in Spanish; ‘to remember’ and ‘to wake up’ in Portuguese), the examples have a different and ambiguous meaning if we take into account that they were produced by a L1-Portuguese speaker. In the examples, did the speaker not write since he remembered to do so too late, or because he woke up late? In sentence (4) a. *We are not obliged to bear* [Spanish meaning] / *support* [L1 meaning] *the other’s decision* both interpretations also could be acceptable. The verb refers to a negative decision, i.e., smoking in front of non-smokers, and as such the use of *bear* would be more reasonable, but at the same time the expression *apoyar decisiones/support decisions* is common in Spanish and English.

The utterances included in (4–7) b. are also in principle correct in Spanish (notwithstanding their grammatical and orthographical errors), but their meaning might be unusual, due to the context and the logical thinking of the interlocutor, as in (1–2). In these cases, knowing that they were produced by non-native speakers and contain a false cognate may cause uncertainty in the interlocutor. Furthermore, the interpretation of

these utterances makes more sense considering the meaning of the false cognates in the speakers' L1. For example, in 4) b. *She always bore* [Spanish meaning]/ *supported* [English meaning] *their sons and daughters*, it is more plausible that a mother would support her children rather than bear them, although this would not be the case if she had problem children, for example. The same occurs in 7) b.: *We still have four uncles and five cousins*/ *We have four uncles and five cousins, however* [L1 meaning], *we don't have neither* (. . .). Although the interpretation with the Spanish meaning is correct, given the context and the expectations of the reader, it is surprising that the speaker says that he "still" has four uncles and five cousins, as if he were expecting them not to be alive. In 6) b. the speaker states *I had to be at the airport at 07:00, but that was the time I remembered* [Spanish] *it/ I woke up* [Portuguese]. Once again, it is surprising, but not impossible, that a person should forget that they have to catch a flight, but it seems more likely that they overslept, i.e. the meaning of the verb in the learner's L1.

In short, working with semi-natural data provides more direct and realistic measurements of L2 learners' linguistic practices. It is therefore easier to understand how different interpretations can arise from the interlanguage if we know that some words have a different meaning in the speakers' L1. Nevertheless, there are some limitations on the use of corpus data in experimental research. For instance, only a few utterances containing false cognates that could be ambiguous were found, even when searching in two different corpora. Furthermore, there is no wide range of false cognates, and the utterances retrieved are not homogeneous, as can be seen when comparing examples a. and b. For all these reasons, a questionnaire designed *ad hoc* with no cases of polysemic and homonymic words in Spanish and with clearer and more varied examples would perhaps be more suitable for the experimental task, since it would offer a systematic approach to both the data collection and data analysis. However, the analysis of real samples from corpora gives researchers a better understanding of how these words are indeed used in the interlanguage by non-native speakers, and how different interpretations can arise from them.

5.2. Ambiguity study

Table 2 presents the percentages of ambiguity detected or otherwise by the two groups, and the Spearman's p statistical coefficient for the comparison between each type of group. The critical group knew that the sentences were produced by L2 speakers, and the meaning of those words in their L1; the control group did not have any information about the speakers who produced the sentences or the false cognates.

Spearman's rank correlation coefficient assesses the relationship between two variables. In this case, these variables are the percentages

Table 2. Responses by the participants in each group and Spearman's p values.

	Critical group		Control group		Significance p -value
	Ambiguity	No ambiguity	Ambiguity	No ambiguity	
<i>Atender</i>	58%	42%	23.8%	76.2%	** .001
<i>Soportar</i>	48%	52%	21.4%	78.6%	* .008
<i>Todavía</i>	60%	40%	11.9%	88.1%	** .002
<i>Acordar</i>	52%	48%	14.3%	85.7%	** .001
<i>Banco</i> (distractor)	52%	48%	47.6%	52.4%	.980
<i>Solo</i> (distractor)	64%	36%	57.1%	42.9%	.158
Average	55.6%	44.4%	29.4%	70.6%	** .001

of ambiguity detected and the type of group. The results confirm that the participants who knew that these sentences were produced by non-native speakers and the meaning of the words in their native languages found more instances of ambiguity. Indeed, the average for the cases of ambiguity (including distractors) is 55.6% in the critical group and 29.4% in the control groups respectively. Additionally, with the exception of distractors, the differences between the two groups are statistically significant in all cases ($p \leq 0.05$). In this context, distractors show a clear indication of the influence of the prior information about the speakers. While such different perceptions are found between the critical and the control groups in all the target items, the results for the distractor items are quite similar (52% and 47.6% of ambiguity detected respectively). There is therefore no statistical significance in those cases, in contrast to the other utterances.

Surprisingly, evidences of lexical ambiguity are detected in all sentences by both groups. The control group was not anticipated to find ambiguity in these utterances (except in those containing distractors) which apparently have only one interpretation. The reason for this pattern could be the fact that as mentioned above, the false cognates included in the task are also polysemic words in Spanish. Other polysemic and homonymous words in the utterances were also mentioned by the students. This is the case with *primo*, which means 'cousin' in the target sentence, but can also mean 'chump' or 'prime number' in other contexts, as well as *citar*, which means 'to make an appointment' in the item, but can also mean 'to quote'. It is therefore likely that some participants paid more attention to the specific words than to the complete utterances in which they appeared. In any case, the results confirm that knowing that

these sentences were produced by L2 speakers and the meaning of the false cognates in their L1 is significant: the percentages for lexical ambiguity for the word *atender* rise to 58% for the critical group (23.8% in the control group); the percentages for ambiguity for the word *todavía* are 60% in the critical group and 11.9% in the control group, etc. Further analyses to confirm whether these results are also connected with the speakers' perceptions of their correctness are provided in the next section.

5.3. Correctness assessment

Table 3 presents the scores displayed in each sentence by the critical and the control groups, ranging from 0 (completely wrong sentence) to 3 (completely correct sentence). The results indicate that both groups present similar patterns to those observed in the ambiguity study. A significant group main effect was found, in which the sentences scored by the control group had better overall results than those scored by the critical group:

Table 3. Scores for utterances by the total participants in each group and Spearman's p values.

	Critical group	Control group	Significance (p -value)
<i>Atender</i>	1.46	2.1	*.005
<i>Soportar</i>	1.98	2.36	.139
<i>Todavía</i>	1.72	1.9	.573
<i>Acordar</i>	1.44	2.38	** .002
<i>Banco</i> (distractor)	2.32	2.28	.861
<i>Solo</i> (distractor)	2.32	2.36	.837
Average (without distractors)	1.87	2.25	** .001

The participants in the critical group found more instances of lexical ambiguity, and also considered that the sentences were less correct. As can be observed in Table 3, the scores are higher in the control group than in the critical group. Moreover, while the average (with the exception of the distractors) is 2.2 in the control group, it is only 1.6 in the critical group. A Spearman correlation was carried out on the mean score that participants awarded the sentences. The results show a significant main effect on the different groups when assessing the correctness of the utterances, $\rho = .161$, $p < .001$, which indicates that participants were influenced when they knew that the sentences were

produced by non-native speakers. However, in some specific cases these differences between the two groups are not statistically significant (e.g. *soportar*, *todavía* and the distractor items).

Interestingly, distractors obtain the best scores in the critical group, and are higher than the average (2.32/3 vs. 1.87/3), while in the control group they are slightly above the average (2.28/3, 2.36/3 vs. 2.25/3). These results show that the participants in the critical group are strongly influenced by knowing that the speakers are non-native, in that they found more instances of lexical ambiguity and awarded a lower score for correctness to the sentences produced by non-native speakers than they did to the distractors. In this context, they found the distractors more correct, confirming the previous hypothesis.

Taking into account the differences between utterances a. and b. set out in 4.1., Figures 1 and 2 show the specific distribution of scores for each sentence provided by both the critical and the control groups.

These figures show that most of the scores awarded by the control group are between 2 and 3, whereas the results from the critical group are between 1 and 2. There does not seem to be a major difference in the control group when assessing sentences a. and b. Indeed, the only major difference is in *todavía* a. vs. b. The other items do not show any tendency, which means that although the utterances of b. seemed to be unusual and a contradiction of logical thinking, these do not affect the correctness and the score awarded. However, this is not the case for the critical group: all the unusual sentences are less correct for these participants, as can be observed in Figure 1. The most evident case is *atender*, for which the first utterance was scored between 2 and 3 by 64% of participants, while the second one only received 24%. Meanwhile, *acordar* has more similar results, with 56% for utterance a. vs. 52% for utterance b.

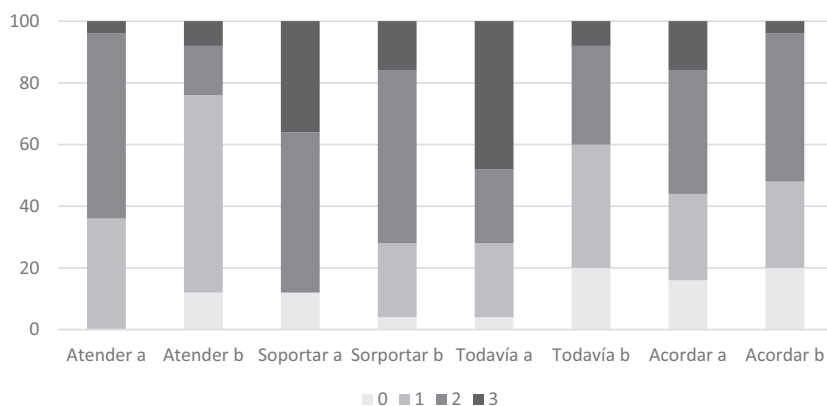


Figure 1. Scores awarded by the critical group.

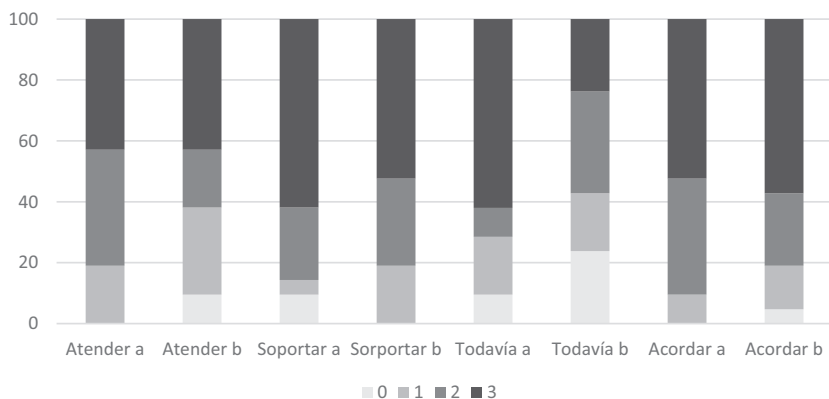


Figure 2. Scores awarded by the control group.

In overall terms, the most correct sentences for the control group are those that contain the words *acordar* and *soportar*. The sentences with *soportar* were also scored highest by the critical group. For these latter participants, the least correct utterances were surprisingly those with the verb *acordar*. While these data do not seem very impressive, they do show that on average, speakers do not use the same criteria when assessing the same utterances: the participants seem to be negatively influenced when they know that sentences are produced by non-native speakers. In those cases, they are more critical towards these speakers, especially if the utterances that they assess do not seem to follow a logic.

6. Conclusion

This exploratory study aimed to shed new light on the specific aspects of false cognates. From a broader perspective, it has also analyzed whether there are any differences between the way that native speakers interpret a type of Spanish utterance when they know or do not know that they have been produced by non-native speakers. The results of a comparative study based on data provided by CAES and CEDEL2 corpora show that this information significantly affects the interlocutor's perception of the message. In this regard, the participants proved to be more critical when they knew that the utterances were produced by a non-native speaker, and especially if these utterances were somewhat unusual or illogical. Furthermore, they were more inclined to find lexical ambiguity in those sentences than participants who did not have this information. These different interpretations could have been due to the orthographical, grammatical, communicative or pragmatic errors of the interlanguage, but these sentences were corrected beforehand in order to avoid response biases or influencing the participants. The results therefore show that in

order to assess utterances, participants pay more attention to the speakers' profiles than to the utterances at the discourse level, even if the sentences do not contain any errors. The interlanguage consequently seems to be subject to a more negative perception than the L1.

Despite the contributions of the present study, some limitations need to be acknowledged. First, this study is based on semi-natural data provided by learners' corpora, which did not include many utterances containing a false cognate that had more than one meaning. Furthermore, examples highlighting lexical ambiguity were limited and not very varied. It is therefore likely that different interpretations could be provided if the study had been based on more situations created *ad hoc* with more diverse utterances. For these reasons, it might be useful to replicate the findings of this study with controlled data and a larger quantity of realistic situations.

In short, this exploratory study is a first step towards going beyond the focus on the false cognates from a theoretical perspective or based on the difficulties that non-native learners encounter when learning the distinction between the meaning in the L1 and in the Spanish L2. Nevertheless, more sources of data will be necessary in order to provide a full picture of the lexical ambiguity in L2 and native speakers' perception of the interlanguage.

Acknowledgement

This study was funded through a grant from Xunta de Galicia (Spain) titled "Use of general and learner corpora for the research and teaching of discourse in Spanish as a foreign language" [grant code: ED481D-2022-016].

References

- ALBA QUIÑONES, V. 2009. El análisis de errores en el campo del español como lengua extranjera. Algunas cuestiones metodológicas. *Revista Nebrija de Lingüística aplicada a la enseñanza de las lenguas* 5:1-16.
- BELDA TORRIJOS, M. 2015. Errores del uso del español en estudiantes extranjeros: creación de una taxonomía y su etiquetado. PhD diss.
- CAES. *Corpus de Aprendices de Español*. Universidade de Santiago de Compostela & Instituto Cervantes [on-line]: <https://galvan.usc.es/caes/>
- CEDEL2. *Corpus Escrito del Español-L2*. Universidad de Granada [on-line]: <http://cedel2.learnercorpora.com/>
- CORDER, S. P. 1967. The Significance of Learners' Errors. *International Review of Applied Linguistics in Language Teaching* 5:161-170.
- COUNCIL OF EUROPE, 2001. *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: Cambridge University Press.
- FERNÁNDEZ SOLERA, S. 2002. Factores sintácticos y semánticos en el procesamiento del lenguaje. PhD diss.

- HAMILTON, R. P. 2001. The insignificance of learners' errors: a philosophical investigation of the interlanguage hypothesis. *Language and communication* 21:73–88.
- HORNO, M. & LÓPEZ, N. 2020. Polisemia y homonimia en la enseñanza no universitaria: revisión y propuesta didáctica. *REGROC: Revista de gramática orientada a las competencias* 3:37–54.
- LARSEN-FREEMAN, D. 2011. *Techniques and Principles in Language Teaching*. Oxford: Oxford University Press.
- LOZANO, C. 2022. CEDEL2: Design, compilation and web interface of an online corpus for L2 Spanish acquisition research. *Second Language Research* 38.4:965–983.
- NAVARRO GALA, R. 1999. Análisis de errores en el marco de la lingüística contrastiva: la expresión escrita. *X Congreso Internacional de la ASELE: Nuevas Perspectivas en la Enseñanza del Español como Lengua Extranjera*. eds. M. Franco, C. Soler, Javier de Cos et al., 481–490. Cádiz: Universidad de Cádiz.
- PENADÉS MARTÍNEZ, I. 2003. Las clasificaciones de errores lingüísticos en el marco del análisis de errores. *Linred: Lingüística en la Red 1*: <https://ebuah.uah.es/dspace/handle/10017/24703>.
- RAMOS SAÑUDO, A. M. 2021. Los marcadores discursivos de cierre en el español como lengua extranjera. *Aportes del CAES a la enseñanza del español como lengua extranjera: estudios aplicados de gramática y pragmática*. eds. C. Ballesterero de Celis & M. Sampedro Mella, 257–285. *Verba: Anuario galego de filoloxía* 81. Universidade de Santiago de Compostela.
- ROCA VARELA, M. L. 2010. The problem of false friends in learner language: Evidence from two learner corpora. *Language Windowing through Corpora*. eds. I. Moskowich, B. Crespo, I. Lareo Martín & P. Lojo Sandino. Coruña: Universidade da Coruña.
- ROJO, G. & PALACIOS, I. 2016. Learner Spanish on Computer. The CAES 'Corpus de Aprendices de Español' Project. In M. Alonso-Ramos (ed.), *Spanish Learner Corpus Research. Current Trends and Future Perspectives*. Amsterdam: John Benjamins, 55–87.
- SAMPEDRO MELLA, M. & SÁNCHEZ GUTIÉRREZ, C. H. 2017. Efecto de la enseñanza extensiva en la memorización del léxico: un estudio de investigación-acción. *Revista Sintagma* 29:61–78.
- SAMPEDRO MELLA, M. 2021a. La expresión del ofrecimiento y la petición en aprendices de español de L1 portugués. *Círculo de lingüística aplicada a la comunicación* 85:113–125.
- SAMPEDRO MELLA, M. 2021b. *Estimado Sr.* vs. *Hola hotel*: el análisis contrastivo de la interlengua para la enseñanza de la variación pragmático-discursiva. *La variación en español y su enseñanza. Reflexiones y propuestas didácticas*. eds. V. Marcet, V. Álvarez Rosa & M. Nevot, 134–150. Salamanca: Universidad de Salamanca.
- SHAFFER, D.E. 2005. Classifying Language Learning Errors. *The Internet TEFL Journal* 58:1–7.
- SELINKER, L. 1972. Interlanguage. *International Review of Applied Linguistics* 10:209–241.
- STAMENOV, M. 2009. *Cognates in language, in the mind and in a prompting dictionary for translation*. Cognates in language, in the mind and in a prompting dictionary for translation. eds. S. Göpferich, A. Lykke Jakobsen, I. M. Mees, 219–252. Denmark: Samfundslitteratur.
- THOMAS, J. 1983. Cross-cultural pragmatic failure. *Applied Linguistics* 4:91–112.
- ULLMANN S. 1976. *Semántica. Introducción a la ciencia del significado*. Madrid: Aguilar.

20 María Sampedro Mella

VALVERDE MATEOS, A. 2012. Análisis de errores de aprendientes de francés lengua extranjera (FLE) basado en corpus orales. PhD diss.

VÁZQUEZ, G. E. 1991. *Análisis de errores y aprendizaje de español / lengua extranjera*. Frankfurt am Main: Peter Lang.

WANG, D. 2021. An Analysis of Tolerance of Ambiguity in Second Language Learning. *Sino-US English Teaching* 18.9:250–256.

Received September 13, 2023

Accepted September 28, 2023

*María Sampedro Mella
Facultad de Filología
Avda. Castelao, s/n,
15705, Santiago de Compostela
Spain
maria.sampedro.mella@usc.es*