

The synchronous and asynchronous learning of anaphora: A corpus-based analysis with learners of English and Spanish

Amanda Maraschin Bruscato – Jorge Baptista
University of Algarve / Portugal

Abstract – This paper aims to investigate the use of nominal, pronominal, and zero anaphora among native speakers of Brazilian learning Spanish or English. To this purpose, two learner corpora were employed: the *Brazilian Learners of Anaphora in English* (BRANEN) and the *Aprendices Brasileños de Anáfora en Español* (BRANES). Participants were undergraduate students with an intermediate-to-advanced proficiency level in the foreign language (English or Spanish) and were randomly assigned into three groups: one had synchronous lessons on the topic, one had asynchronous lessons, and a third one was the control group (which had no lesson). They all completed short narratives in four moments, and their written texts were compiled to investigate how a different instructional mode can better contribute to the learning of this specific discourse mechanism. Third-person human subjects of finite clauses and their antecedents were manually annotated on *Recogito*. When analysing the pre-test, we found that learners could be less redundant and could use more zero anaphora than pronominal anaphora in English coordinate clauses and Spanish main clauses to continue the topic/subject. The experimental groups practised it during the online course and the asynchronous instructional mode proved to be more effective until the third test (immediately after the course), but the same was not found on the delayed post-test (one month later).

Keywords – anaphora; language learning; asynchronous learning; synchronous learning; learner corpus

1. INTRODUCTION

The processing of anaphora has been the focus of many studies over the past years, due to the importance of the mechanism for a cohesive communication in any language. In simple terms, anaphora can be defined as a discourse mechanism in which an element in the text (anaphor) refers back to another element (antecedent), as in (1), where the pronouns *she* and *her* refer back to *Mary*.

(1) **Mary** fell. **She** was still on the ground when Peter found **her**.



In the previous example, there was only one possible antecedent, but in discourse there are usually more possibilities. Although the human brain is generally able to identify the correct antecedents, natural language processing has been a challenge to computational linguistics, due to the difficulty of training a machine to understand how the human brain works. Thus, many researchers have been trying to comprehend better how speakers of different languages correlate anaphors and antecedents.

As Lozano (2021a) suggests, research on anaphora resolution is relevant to investigate cross-linguistic influence and second language (L2) development. Many studies have analysed anaphora resolution through questionnaires with ambiguous sentences or through learner corpora, as one of the works presented in the first international conference on *The Acquisition and Processing of Reference and Anaphora Resolution* (APRAR), organised by the Vrije Universiteit Brussel (Belgium) and the University of Granada (Spain) in 2021. Lozano's research with the *Corpus Escrito del Español L2* (CEDEL2; Lozano 2016, 2021b),¹ for example, focused on how learners of different first languages (L1) produce anaphora in Spanish. However, although these studies consider different language proficiencies, they do not have a pedagogical approach.

Our study contributes to the field by presenting and analysing two new learner corpora: the *Brazilian Learners of Anaphora in English* (BRANEN) and the *Aprendices Brasileños de Anáfora en Español* (BRANES), built to investigate the learning of anaphora under two instructional modes (synchronous and asynchronous). The novelty of this research is to investigate anaphora resolution in a combination of languages that has not been the focus of most studies so far (Brazilian Portuguese L1 and English or Spanish L2), and to collect written data at four points in time (pre-test, post-test 1, post-test 2, delayed post-test) to investigate the learning of nominal, pronominal, and zero anaphora during an online course focused on this mechanism.

A total of 45 participants, who were Brazilian undergraduate students, had an intermediate-to-advanced proficiency level in the foreign language (30 students of English and 15 of Spanish) and were randomly assigned into three groups: one had synchronous lessons on the topic, one had asynchronous lessons, and a third one was the control group (which had no lesson). They all completed short narratives in four moments,

¹ <http://cedel2.learnercorpora.com/search>

and their written texts were compiled to investigate how a different instructional mode can better contribute to the learning of this specific discourse mechanism.

By the analysis of these corpora, this paper intends to answer the following research questions: (RQ1) Are the new corpora representative of Brazilian learners of English or Spanish? (RQ2) What are the differences between Brazilian learners of English or Spanish on the production of anaphora? (RQ3) What are the differences between the instructional modes (synchronous, asynchronous, and control) on the learning of anaphora?

The outline of the paper is as follows. In Section 2 we discuss the cross-linguistic influence on anaphora resolution, some of the learner corpora available to this purpose, and the effects of instructional modes on the L2 learning of anaphora. In Section 3 we describe the research method of this study and in Section 4 we present our findings and the analysis of the results. The paper closes with some considerations for future investigations.

2. REVIEW OF LITERATURE

2.1. Cross-linguistic influence on anaphora resolution

As previously explained, anaphora can be defined as a discourse mechanism in which an element in the text (anaphor) refers back to another element (antecedent). There are many types of anaphoric elements, such as nominal, pronominal, and zero anaphora, and their use differs across languages. Pronominal anaphora, for example, is predominant in English, a non-null-subject language, while zero anaphora prevails in Portuguese and Spanish, known as null-subject languages (Chomsky 1981; Rizzi 1982). Still, zero anaphora is commonly used in English in coordinate clauses with the same subject (Quesada and Lozano 2020), something which English learners might overlook when writing in the foreign language.

The anaphoric setup in the learners' L1 can influence their anaphoric behaviour in the L2 and, to avoid ambiguities and misunderstandings, L2 intermediate learners tend to be more explicit than native speakers in their discourse (Hendriks 2003). Although the amount of over-explicitation varies according to the target language, the preference to be redundant rather than ambiguous is related to pragmatic felicity (Lozano 2016). Considering Grice's (1975) second maxim of Quantity (do not make your contribution

more informative than is required), speakers should use null forms whenever possible, as in (2); and, according to the second maxim of Manner (avoid ambiguity), they should prefer redundancy over ambiguity, as in (3).

(2) **Mary** arrived home and **she/Ø** called Anna.

(3) Mary called **Anna** when **she/Anna** was travelling.

Focusing on the learning of Spanish by English native speakers, Lozano (2016) analysed the *Corpus Escrito del Español L2* (CEDEL2) and proposed the *Pragmatic Principles Violation Hypothesis*. According to it, very advanced Spanish learners prefer to be redundant by using pronouns to continue a topic, as in (4), than to be ambiguous by omitting the subject when changing it, as in (5). The author also suggests that L1 and L2 speakers tend to use nominal instead of pronominal anaphora to avoid ambiguity when there are same-gender antecedents.

(4) **María** nos llamó cuando **ella** estaba viajando. (redundant) ‘**Mary** called us when **she** was travelling’.

(5) María llamó a **Ana** cuando **Ø** estaba viajando. (ambiguous) ‘**Mary** called **Anna** when **Ø** was travelling’.

As Miltsakaki (2002) explains, there are many aspects that influence topic continuity and topic shift, some of which are syntactic. In our previous studies (Bruscatto and Baptista 2021d, 2022a, 2022b), we tested different anaphoric strategies used by Portuguese, English, and Spanish learners and native speakers when reading ambiguous sentences. We found out that, while English and Spanish native speakers interpret subject pronouns as continuing the topic, Portuguese native speakers interpret them as topic shifters.

Whereas our previous studies analysed data collected with reading questionnaires, the current paper provides more information on the topic by using corpora to analyse written data. The present investigation aims to answer whether Brazilian learners of English or Spanish are more redundant or ambiguous in their own texts and how they learn to reduce these issues over-time.

2.2. *The effects of instructional modes on the L2 learning of anaphora*

Several studies have investigated learners’ knowledge of anaphora, as shown by Ellis (2008: 608–609), and many others have investigated the impact of the instructional mode

on language learning, as reported by Siemens *et al.* (2015). However, as Li (2014) explains, there is practically no research connecting both topics. Liu (2010), for example, investigated whether the type of feedback (implicit or explicit) used in Computer Assisted Language Learning (CALL) would have any impact on the learning of pronominal anaphora in English as L2. The researcher prepared computer exercises on anaphora and asked 28 Chinese adults with an intermediate level of English proficiency to answer them twice. Half of the group read an explanation after each error, while the rest just received a right or wrong feedback. In the end, there was no difference between the groups, probably because they did not have any lessons, the exercises only took half an hour, and were the same both times.

We only found one previous study that investigated the impact of the instructional mode on the learning of anaphora. Li (2014) compared the learning of zero anaphora in Chinese by English native speakers who had onsite or online lessons on the topic, but the effect of the instructional mode to learners' text production was not investigated. Still, the results showed that those who had asynchronous lessons performed better than the others. Possibly an asynchronous course could also improve students' writing, something the current paper intends to answer.

2.3. Learner corpora available

There are many corpora available to study learners' production, such as the *International Corpus of Learner English* (Granger 2003), the *Multilingual Learner Corpus* (Tagnin 2006), the *Corpus of English as a Foreign Language* (COREFL; Lozano *et al.* 2020), and CEDEL2 (Lozano 2021b), as well as some native corpora built to investigate specific types of anaphora, such as *OntoNotes* (Pradhan *et al.* 2007), *Anaphora Resolution and Underspecification* (Poesio and Artstein 2008), and *WikiCoref* (Ghaddar and Langlais 2016). In Portuguese, for example, there are corpora focused on zero (Baptista *et al.* 2016), pronominal (Marques 2013), or nominal anaphora (Pardo *et al.* 2017). However, these are native corpora, and no learner corpus seemed to have the instructional mode as a variable before BRANEN and BRANES.

An example of a multilingual learner corpus that compiled written synchronous and asynchronous computer-mediated communication texts is the *Multilingual Learner Corpus* (MiLC; Andreu *et al.* 2010). However, it was only used to investigate

interlanguage errors in teleconferences and emails, and does not take into account the learning progress in online instructional modes. BRANEN and BRANES, therefore, bring a new perspective to corpora studies by considering different instructional modes in their compilation and analysis.

3. METHOD

3.1. Participants

In this paper we present two corpora, BRANEN and BRANES, which contain texts written by 45 Brazilian undergraduate students with a major in English or Spanish who were in the third or fifth semester of their courses. Texts were collected during a short online course on anaphora in the first semester of 2020.

There were 15 Spanish learners and 30 English learners. Most of them (62%) were in the third semester of their courses, 73 per cent were female, and the average age was 20 (they were between 18 and 41 years). For each language, participants were randomly divided into three groups: one had two synchronous lessons on anaphora, another had two asynchronous lessons, and the control group did not take any lessons.

All participants agreed to take part in the research and answered a grammar questionnaire to ensure they had an intermediate-to-advanced proficiency level in the foreign language. The proficiency test had 20 reading questions, taken from Cambridge University or the Cervantes Institute, equally distributed between levels A2 and C1. Although the English learners' scores were a bit under the Spanish learners' scores, the results among groups were very similar, as can be seen in Table 1 below.

| Language | Group | Mean | Standard Deviation |
|----------|---------------------|------|--------------------|
| Spanish | Synchronous (N=5) | 15 | 2 |
| | Asynchronous (N=5) | 15 | 2 |
| | Control (N=5) | 14 | 5 |
| English | Synchronous (N=10) | 12 | 4 |
| | Asynchronous (N=10) | 14 | 3 |
| | Control (N=10) | 13 | 5 |

Table 1: Grammar test results (retrieved from Bruscatto and Baptista 2021c)

3.2. *Experiment*

The university e-learning platform (Moodle) was used during the course. The synchronous groups participated in two videoconference lectures of 90 minutes each, while the asynchronous groups watched short videos, read texts, participated in discussion forums, and answered automatic exercises. As explained in Bruscato and Baptista (2021c: 7), each experimental lesson included:

activation of prior knowledge on the topic; lecture on anaphora for half an hour; reading and analysis of material; group discussion; reading and writing exercises; and feedback. [...] In the first lesson, students introduced themselves; learned about cohesion; the types of anaphora; and the subject, object, and possessive pronouns in the language of study; worked with corpus; completed sentences with the correct pronouns; and did an exercise similar to the test. In the second lesson, they were challenged to solve the ambiguity of some sentences; learned about ambiguity resolution, demonstrative, and relative pronouns; corrected and completed some sentences with pronouns; analysed the coreferences in a fable, comparing their manual analysis with an automatic one; and, again, they did an exercise like the test.

As shown elsewhere (Bruscato and Baptista 2021a, 2022b, 2022c), students wrote short narratives of 100–150 words in four different moments: before the course started (to check students' performance before the intervention), after the first lesson (to measure their progression during the course), after the second lesson (to check their progression when they completed the course), and a month after the course ended (to investigate if students still remembered what they studied). Texts were different but followed the same structure, they were all third-person narrative fictional texts with multiple antecedents.

After reading the beginning of a story (with ten hidden anaphoric problems to solve, previously analysed in Bruscato and Baptista 2021c), students corrected the mistakes they found, and were then asked to conclude the text and submit their files via Moodle. The task was planned to ensure that every student would write about the same topic and that there were multiple antecedents in the texts. The pre-test is presented below.

Instruction: Read the beginning of a narrative and correct the mistakes you find, then write an end to the story between 100 and 150 words.

John and Mary were twins and they were only twelve years old when became orphans. Before these misfortune, John and Mary lived with them parents, Joseph and Ana, that loved they very much. They were all happy, until the country declared war. Joseph was sent to fight, and his wife had to take care of the children and the house. One day, a letter from the government arrived. Ana already knew her content: hers husband was dead. The widow became herself deeply depressed and could not get out of bed. In despair, John and Mary decided to visit the only neighbour they had (they called her witch) to ask for help.

3.3. Corpora

The corpora were first made available on *Sketch Engine*² (Kilgarriff *et al.* 2014), a corpus managing and text analysis software, and include metadata about the participants' group (asynchronous, synchronous, control) and testing moment (1, 2, 3, 4). The *Sketch Engine* corpus query system was chosen because it is commonly used by linguists, and because the *European Lexicographic Infrastructure*³ project provides all academic institutions in the EU free access to the software, at least until 2022.

After their compilation, the corpora were manually annotated using the *Recogito* annotation tool,⁴ an online free software that allows the establishment of unilateral, oriented relations between anaphors and antecedents. An anaphora expert annotated the whole corpora, while another expert was responsible for annotating 20 per cent of the texts, which were randomly selected. After the annotation was completed, the intercoder reliability coefficients were calculated using *ReCal2: Reliability for 2 Coders*.⁵ The codes used for the anaphors and to establish the anaphoric relation were very similar (agreement around 95%), and the chosen antecedents were the same in about 85 per cent of the time.

Since the first text was written before the course started and we aimed to analyse how learners processed anaphora in comparison to native speakers, three Spanish and six English native speakers also volunteered to do the first task. Their texts were annotated and were made available with the learner corpora.

Based on Lozano's (2016) annotation scheme of subject expressions, third-person human subjects of finite clauses and their antecedents were annotated following the

² www.sketchengine.eu

³ <https://elex.is/>

⁴ <https://recogito.pelagios.org/>

⁵ <http://dfreelon.org/utis/recalfront/recal2/#doc>

scheme shown in Table 2. First, the form of the expression was annotated. Since blank spaces cannot be marked on *Recogito*, in case of zero anaphora the annotation was on the primary verb. All anaphors were subjects, but antecedents could also be non-subjects. Then, they received a tag according to the type of clause they were in, and there was an option to annotate if the expressions were ambiguous or redundant. After that, the intrasentential and intersentential relations were established. When necessary, there was the possibility to specify if it was a case of cataphora or a partial relation.

The annotation scheme with the tags and examples from the corpora are presented in Table 2.⁶

| Form | |
|--------------------------------------|---|
| Zero – <i>zr</i> | She was good, generous, and helped other people [...]. ia1e[1] |
| Pronoun – <i>pp</i> | [...] they would get what they wanted [...]. Ia1c |
| Determiner – <i>dt</i> | [...] the two started [...]. Ia3c |
| Common noun – <i>nc</i> | [...] the twins arrived home late [...]. Ia1c |
| Proper noun – <i>np</i> | Mary knocked on the door [...]. Ia1a |
| Function | |
| Subject – <i>sj</i> | They missed their dad [...]. ia1c |
| Non-subject – <i>ns</i> | The witch told them that they had to [...]. ia1c |
| Clause | |
| Main clause – <i>mc</i> | They went there for help [...]. ia1c |
| Coordinate clause – <i>cc</i> | They missed their dad and they were worried about their mom [...]. ia1c |
| Subordinate clause – <i>sc</i> | The witch told them that they had to [...]. ia1c |
| Pragmatic-felicity (optional) | |
| Ambiguous – <i>am</i> | She also tried to help the kids with their mother, but she ended up very sick [...]. is1d |
| Redundant – <i>rd</i> | They missed their dad and they were worried about their mom [...]. ia1c |
| Relation | |
| Intrasentential – <i>in</i> | They missed their dad and they were worried about their mom [...]. ia1c |
| Intersentential – <i>tr</i> | They went there for help. They wanted their old life back [...]. ia1c |
| Cataphora (optional) – <i>ca</i> | Since she was a kid, Mary knew [...]. ic1j |
| Partial (optional) – <i>pr</i> | She seemed happy to help their mother, so they all went to the bedroom where Anna was [...]. ia1b |

Table 2: Annotation scheme with tags and examples

Possessive, reflexive, and relative pronouns were not annotated in this phase. This study focuses on subject expressions and their antecedents and, since relative pronouns appear right after their antecedent in the text, they were not relevant for the current purposes.

Figures 1 and 2 show examples of annotated texts. As can be seen, the English text had many subject pronouns (e.g. *They went there for help. They wanted their old life*

⁶ The code after each example indicates a specific file in the corpora.

back. *They* missed their dad and *they* were worried about their mom.). The Spanish excerpt, on the contrary, had no subject pronoun, but exhibited many cases of zero and nominal anaphora (e.g. *La bruja* había puesto veneno en la sopa y se *reía*. ‘The **witch** had added poison to the soup and **was laughing**.’). After the annotation was completed, results were exported into .csv files and analysed in SPSS (v. 26; IBM Corporation 2020).

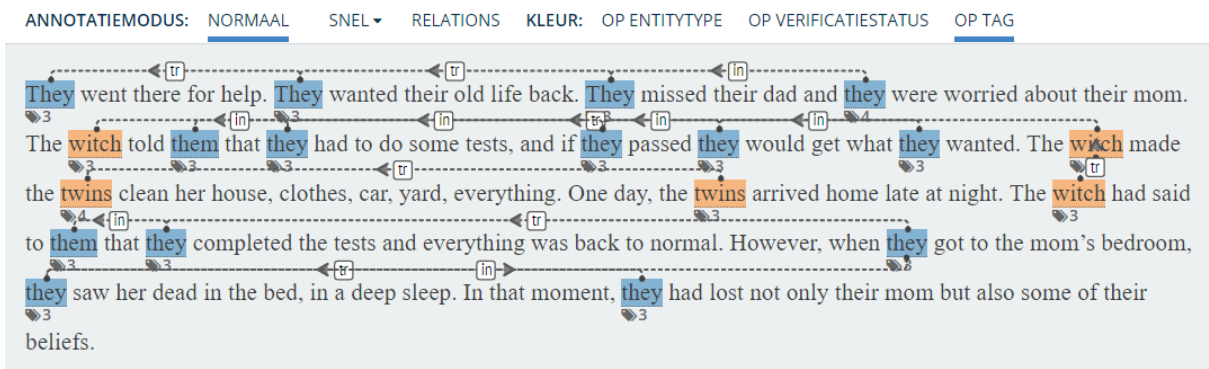


Figure 1: Subject anaphora annotation in English narrative

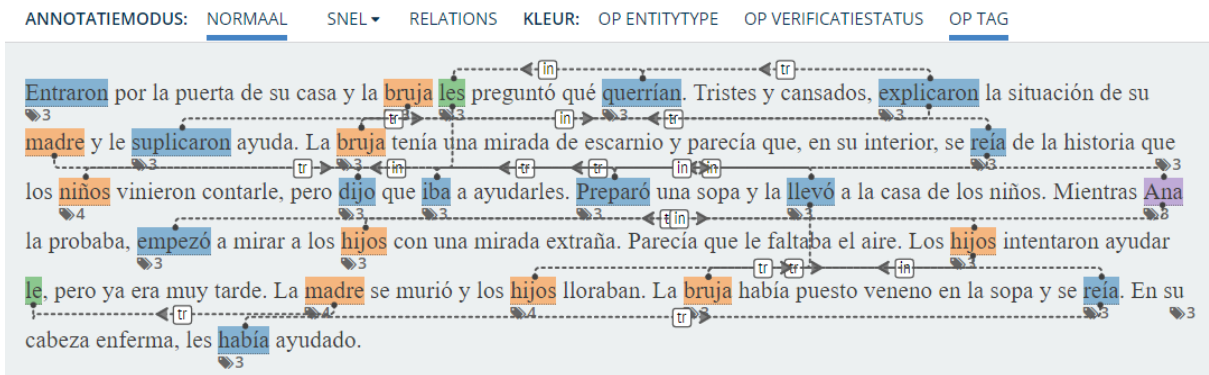


Figure 2: Subject anaphora annotation in Spanish narrative

BRANEN has 120 documents and was automatically part-of-speech (POS) tagged by *Sketch Engine* with the *Modified English TreeTagger*, while BRANES has 60 documents and was POS-tagged with the *Spanish FreeLing* tagset. Table 3 presents the size of the corpora. More information about them can be found in Bruscatto and Baptista (2021c).

| | documents | sentences | lemmas | unique words | words | tokens |
|--------|-----------|-----------|--------|--------------|--------|--------|
| BRANEN | 120 | 1,069 | 1,678 | 2,242 | 17,454 | 19,934 |
| BRANES | 60 | 543 | 1,299 | 2,095 | 9,021 | 10,233 |

Table 3: Size of the corpora (data retrieved from *Sketch Engine*)

4. RESULTS AND DISCUSSION

The results retrieved from the corpus-based analysis will be presented and discussed in what follows. Before that, however, some descriptive information about the anaphoric relations in the corpora and the distribution of the anaphoric forms will be provided. Tables 4 and 5 below present the number of anaphoric relations per group and test moment.

| | BRANEN | BRANES |
|--------------|---------------|---------------|
| Asynchronous | 616 (35.9%) | 310 (33.6%) |
| Control | 516 (30.1%) | 285 (30.9%) |
| Synchronous | 582 (34%) | 327 (35.5%) |
| Total | 1,714 (100%) | 922 (100%) |

Table 4: Number of anaphoric relations per group

| | BRANEN | BRANES |
|--------|---------------|---------------|
| Text 1 | 427 (24.9%) | 234 (25.4%) |
| Text 2 | 397 (23.1%) | 202 (21.9%) |
| Text 3 | 418 (24.4%) | 235 (25.5%) |
| Text 4 | 472 (27.5%) | 251 (27.2%) |
| Total | 1,714 (100%) | 922 (100%) |

Table 5: Number of anaphoric relations per test moment

The differences in the number of anaphoric relations among groups (Table 4) was negligible in both corpora, since there was only around a five per cent difference between the minimum and maximum values. The same was found when considering the test moments (Table 5). Still, the control groups established fewer relations than the others in both corpora. Considering the four test moments, which were done under similar conditions, there were slightly fewer anaphoric relations in the second test and an above-average number of relations in the fourth, but the number of anaphoric relations was highly correlated in the two corpora among groups ($r = 0.918$) and test moments ($r = 0.953$). This was already expected, since all participants were instructed to write a similar number of words (100–150) in each text and were provided with the same number of possible antecedents.

After identifying a similar number of anaphoric relations among groups and test moments, we compared the distribution of the anaphoric forms used by English and Spanish learners and by native speakers in the first test (Table 6).

| | BRANEN | | BRANES | |
|-------------|-----------------|---------------|-----------------|---------------|
| | Learners (n=30) | Natives (n=6) | Learners (n=15) | Natives (n=3) |
| Zero | 42 (9.8%) | 14 (12.7%) | 128 (54.7%) | 22 (56.4%) |
| Determiner | 4 (0.9%) | 1 (0.9%) | 8 (3.4%) | 0 (0%) |
| Pronoun | 233 (54.6%) | 56 (50.9%) | 25 (10.7%) | 1 (2.6%) |
| Common noun | 97 (22.7%) | 21 (19.1%) | 43 (18.4%) | 11 (28.2%) |
| Proper noun | 51 (11.9%) | 18 (16.4%) | 30 (12.8%) | 5 (12.8%) |
| Total | 427 (100%) | 110 (100%) | 234 (100%) | 39 (100%) |

Table 6: Distribution of the anaphoric forms in the first test

The distribution of the anaphoric forms in the first test between learners and native speakers from BRANEN and BRANES was very similar ($r(EN) = 0.988$; $r(ES) = 0.962$). As Table 6 shows, determiners were hardly used as anaphors (0 to 3.4%) and, in each language, there was a preferred type of anaphora. As expected, in English, more than half of the subject anaphors were pronouns (50.9%), while in Spanish ellipsis was preponderant (56.4%). However, L2 learners produced slightly more pronominal anaphora (+3.7% in English and +8.1% in Spanish) than native speakers. In general, Spanish learners also used less nominal anaphora (-9.8%), while English learners used more common nouns as subjects than natives did (+3.6%).

The similarity between learners and native speakers can be explained by the students' intermediate-to-advanced level of proficiency in the language. Still, there were some slight differences between the groups, showing that students could sometimes replace pronouns with other types of anaphors.

4.1. Representativeness

Since BRANES had a small number of informants (15 Spanish learners and three native speakers), we compared our results with data from CEDEL2 (Lozano 2021b), namely in the use of zero and pronominal anaphora by L1 European Portuguese (L2 Spanish and L1 Spanish adults). For this, the Chaplin task was used, in which participants had to narrate a silent Charles Chaplin video clip. This corpus consists of 137 written texts from native speakers and 85 from learners, of which 96.5 per cent had an intermediate-to-advanced level in the L2, that is, a proficiency level similar to that of the subjects in our study.

For the use of pronominal anaphora, we first looked for instances of third-person nominative personal pronouns in CEDEL2, but not a single occurrence was found. We then checked if there was any nominative pronoun in the corpus, and there were two

occurrences of first-person personal pronouns from learners. One of the sentences was [...] *él encuentra un billete que dice “cuidame, yo soy huérfano”* [...] (‘[...] he finds a note that says “take care of me, I am an orphan” [...]'). Irrespective of the occurrence of pronoun *yo* (I), clearly, there was a third-person nominative personal pronoun in this sentence: *él* (he). However, its case had not been annotated. Since we could not automatically distinguish nominative from other types of personal pronouns using the tool, we left this search for further research. However, other studies interested in this can download the corpus from the CEDEL2 website and manually annotate it.

To compare the use of zero and pronominal anaphora, we checked the number of occurrences of a punctuation mark or a conjunction followed by a third-person verb with either an ellipsis in the middle, as in (6), or the lemma *él* (he), as in (7).

(6) **Sigue caminando y pide** a un hombre que lo sujete por un momento [...]. ‘He keeps walking and asks a man to hold him for a moment [...].’

(7) [...] **cuando ella ve** el nene en su cochecito, **ella corre** en dirección a Chaplin [...]. ‘[...] when she sees the baby in his stroller, she runs to Chaplin [...].’

Table 7 compares the number of zero and pronominal anaphora in BRANES’s pre-test (before the intervention) with the frequencies of comparable patterns found in CEDEL2 (for which no intervention took place). Coincidentally, the percentages found of zero and pronominal anaphora compared to their total were identical between the two groups of learners, and extremely similar between the groups of natives. Despite the small number of participants in BRANES, the similar results found in CEDEL2 give some assurance about the remarks made above. In the next subsections, the data in BRANES and BRANEN are detailed and compared in depth.

| | BRANES | | CEDEL2 | |
|-----------------|-----------------|---------------|-----------------|-----------------|
| | Learners (n=15) | Natives (n=3) | Learners (n=85) | Natives (n=137) |
| Zero anaphora | 128 (83.7%) | 22 (96%) | 231 (83.7%) | 1,187 (97.5%) |
| Pronominal ana. | 25 (16.3%) | 1 (4%) | 45 (16.3%) | 30 (2.5%) |
| Total | 153 (100%) | 23 (100%) | 276 (100%) | 1,217 (100%) |

Table 7: Zero and pronominal anaphora in BRANES and CEDEL2

4.2. Differences between Brazilian learners of English or Spanish on the production of anaphora

To answer the second research question, we compare how participants produced anaphora in the first test (pre-intervention). Table 8 presents the frequencies of the pre-test and shows that, in BRANEN and BRANES, both learners and native speakers established anaphoric relations using the different strategies in a similar way ($r(\text{EN}) = 0.994$; $r(\text{ES}) = 0.957$). This result was already expected, due to the students' proficiency in the language.

| | | BRANEN | | BRANES | |
|------------------------|-----------------|--------------------|------------------|--------------------|------------------|
| | | Learners (n=30) | Natives (n=6) | Learners (n=15) | Natives (n=3) |
| Anaphor clause | Main | 195 (45.7%) | 59 (53.6%) | 104 (44.4%) | 13 (33.3%) |
| | Coordinate | 115 (26.9%) | 27 (24.6%) | 63 (27%) | 15 (38.5%) |
| | Subordinate | 117 (27.4%) | 24 (21.8%) | 67 (28.6%) | 11 (28.2%) |
| Anaphor pragmatics | No problem | 391 (91.6%) | 110 (100%) | 176 (75.2%) | 34 (87.2%) |
| | Ambiguous | 9 (2.1%) | 0 (0%) | 6 (2.6%) | 0 (0%) |
| | Redundant | 27 (6.3%) | 0 (0%) | 52 (22.2%) | 5 (12.8%) |
| Anaphoric relation | Intrasentential | 176 (41.2%) | 40 (36.4%) | 107 (45.7%) | 19 (48.7%) |
| | Intersentential | 251 (58.8%) | 70 (63.6%) | 127 (54.3%) | 20 (51.3%) |
| Antecedent function | Subject | 313 (73.3%) | 86 (78.2%) | 161 (68.8%) | 31 (79.5%) |
| | Non subject | 114 (26.7%) | 24 (21.8%) | 73 (31.2%) | 8 (20.5%) |

Table 8: Frequencies of the first test

Learners behaved almost the same, despite the target language. In general, the differences in their results are less than 5 per cent. Nonetheless, compared to native speakers, some distinctions were found. In learners' texts, almost half of the subject anaphors were in main clauses (46.7% EN; 44.4% ES). In comparison, native English speakers used 7.9 per cent more subject anaphors in main clauses and Spanish speakers used 11.1 per cent fewer subject anaphors in main clauses. In English, there was almost no difference in coordinate clauses, but, in subordinate clauses, natives produced slightly fewer subject anaphors (-5.6%). In Spanish, on the other hand, the results from subordinate clauses were very similar, but, in coordinate clauses, native speakers used 11.5 per cent more subject anaphors than learners did.

Possibly, learners were more influenced by their L1 syntax than by the L2 when using the sentences and therefore behaved the same despite the target language regarding the anaphor's clauses. This is in line with Bruscatto and Baptista (2021d, 2022a, 2022b) regarding the anaphoric strategies used by learners when reading. Based on these results,

it was also found that the preferences in English and Spanish as L1 regarding the distribution of anaphoric subjects in the types of clauses seem to differ. While, for example, in English 53.6 per cent of the subject anaphors were produced in main clauses, in Spanish that number decreased to 33.3 per cent. In further research, it would be relevant to compare these results with data from more informants.

Another difference among the English and Spanish texts is the distance between the anaphors and their antecedents. Although learners and native speakers from each language behaved similarly, English native speakers showed a clearer preference to retrieve intersentential antecedents (63.6%, compared to 51.3% in Spanish). This could be related to the previously discussed higher number of anaphoric subjects in English main clauses.

Besides the preference to select intersentential antecedents, most of them were also subjects among native speakers of English. Although the tendency for topic continuity was already expected, Spanish learners chose a subject antecedent 10.7 per cent less frequently than native speakers and, considering all types of anaphora, they were also 9.4 per cent more redundant. 35 out of their 52 occurrences of redundancy were subjects in a main clause, and nine of these were nouns that retrieved the subject from another sentence, as in (8). Considering the other groups (English learners and all natives), there was not much redundancy in general and even less ambiguity.

- (8) Los **niños**, que también se apegaron a la vecina, muy agradecidos, aceptaron la propuesta. Y aunque tristes, los **niños** estaban muy agradecidos por la compasión y la empatía de su vecina [...] (ec1b). ‘The **children**, who also attached themselves to the neighbour, very grateful, accepted the proposal. And although sad, the **children** were very grateful for the compassion and empathy of their neighbour [...]’

Since a substantial part of the anaphors found in the corpus recover non-subject antecedents, these values call for further analysis. This is the main purpose of Tables 9, 10, and 11, which present the results per group for main, coordinate, and subordinate clauses, respectively.

Table 9 shows the results for anaphora in main clauses. As expected, in English, either nominal or pronominal subjects are used in main clauses. For both groups of informants, there were around 33 per cent of nouns and 39 per cent of pronouns that recovered a subject antecedent. Since English is not a null-subject language, zero

anaphora in subject position of main clauses is not grammatical, and learners complied with this general rule.

| Anaphor form | Antecedent function | BRANEN | | BRANES | |
|--------------|---------------------|--------------------|------------------|--------------------|------------------|
| | | Learners (n=30) | Natives (n=6) | Learners (n=15) | Natives (n=3) |
| Zero | Subject | 0 (0%) | 0 (0%) | 34 (32.7%) | 5 (38.5%) |
| | Non-subject | 0 (0%) | 0 (0%) | 2 (1.9%) | 0 (0%) |
| Determiner | Subject | 2 (1%) | 0 (0%) | 2 (1.9%) | 0 (0%) |
| | Non-subject | 1 (0.5%) | 0 (0%) | 2 (1.9%) | 0 (0%) |
| Pronoun | Subject | 76 (39%) | 23 (39%) | 9 (8.7%) | 0 (0%) |
| | Non-subject | 19 (9.7%) | 7 (11.9%) | 4 (3.9%) | 0 (0%) |
| Noun | Subject | 62 (31.8%) | 21 (35.6%) | 23 (22.1%) | 6 (46.1%) |
| | Non-subject | 35 (18%) | 8 (13.5%) | 28 (26.9%) | 2 (15.4%) |
| Total | | 195 (100%) | 59 (100%) | 104 (100%) | 13 (100%) |

Table 9: Anaphora in main clauses

In Spanish, either nominal or zero anaphora are used. While practically all ellipses recovered a previous subject, nouns were used to recover both, subject and non-subject antecedents. Around a fifth of learners' subject anaphors were nouns that retrieved a previous subject and, as mentioned before, nine of these were redundant. Unlike native speakers, Spanish learners produced pronominal anaphora (but to a lesser extent than English learners).

Table 10 presents the results for coordinate clauses. In coordinate clauses, both languages retrieve intrasentential subjects by zero anaphora. However, English native speakers used 15.4 per cent more ellipses than learners did, while Spanish learners used it 16.2 per cent more frequently than native speakers. Pronominal anaphora was also common in English for the same task, especially for learners who, as already mentioned, used fewer ellipses. To select intersentential antecedents, all groups mostly used nominal anaphora, as in example (8), above.

| Anaphor form | Anaphoric relation | Antecedent function | BRANEN | | BRANES | | |
|--------------|--------------------|---------------------|--------------------|------------------|--------------------|------------------|--------|
| | | | Learners (n=30) | Natives (n=6) | Learners (n=15) | Natives (n=3) | |
| Zero | Intra. | Subject | 42 (36.5%) | 14 (51.9%) | 48 (76.2%) | 9 (60%) | |
| | | Non-subject | 0 (0%) | 0 (0%) | 2 (3.2%) | 2 (13.3%) | |
| | Inter. | Subject | 0 (0%) | 0 (0%) | 2 (3.2%) | 0 (0%) | |
| Determiner | Intra. | Non-subject | 0 (0%) | 1 (3.7%) | 2 (3.2%) | 0 (0%) | |
| Pronoun | Intra. | Subject | 27 (23.5%) | 5 (18.5%) | 1 (1.5%) | 0 (0%) | |
| | | Non-subject | 16 (13.9%) | 2 (7.4%) | 2 (3.2%) | 0 (0%) | |
| | Inter. | Subject | 4 (3.5%) | 1 (3.7%) | 0 (0%) | 0 (0%) | |
| | | Non-subject | 1 (0.8%) | 0 (0%) | 0 (0%) | 0 (0%) | |
| | Noun | Intra. | Subject | 4 (3.5%) | 0 (0%) | 0 (0%) | 0 (0%) |
| | | | Non-subject | 4 (3.5%) | 1 (3.7%) | 0 (0%) | 0 (0%) |
| Inter. | | Subject | 13 (11.3%) | 3 (11.1%) | 4 (6.3%) | 3 (20%) | |
| | | Non-subject | 4 (3.5%) | 0 (0%) | 2 (3.2%) | 1 (6.7%) | |
| | Total | | 115 (100%) | 27 (100%) | 63 (100%) | 15 (100%) | |

Table 10: Anaphora in coordinate clauses

Finally, the results for subordinate clauses are shown in Table 11. In subordinate clauses, there is also a preference to retrieve a subject antecedent, which is usually intrasentential. While Spanish speakers prefer to use zero anaphora, English speakers mainly use pronominal anaphora for that matter. However, there was some difference between the English groups. Natives chose pronouns to recover 11.5 per cent more intrasentential subjects than learners, who chose them to select 9.6 per cent more intrasentential non-subjects. It seems there is a stronger preference for topic continuity in English as L1. Lastly, as already seen before, nominal anaphora tends to select intersentential antecedents.

| Anaphor form | Anaphoric relation | Antecedent function | BRANEN | | BRANES | |
|--------------|--------------------|---------------------|--------------------|------------------|--------------------|------------------|
| | | | Learners (n=30) | Natives (n=6) | Learners (n=15) | Natives (n=3) |
| Zero | Intra. | Subject | 0 (0%) | 0 (0%) | 22 (32.8%) | 3 (27.2%) |
| | | Non-subject | 0 (0%) | 0 (0%) | 10 (14.9%) | 1 (9.1%) |
| | Inter. | Subject | 0 (0%) | 0 (0%) | 7 (10.4%) | 2 (18.2%) |
| | | Non-subject | 0 (0%) | 0 (0%) | 1 (1.5%) | 0 (0%) |
| Determiner | Inter. | Subject | 1 (0.9%) | 0 (0%) | 1 (1.5%) | 0 (0%) |
| | | Non-subject | 0 (0%) | 0 (0%) | 1 (1.5%) | 0 (0%) |
| Pronoun | Intra. | Subject | 45 (38.5%) | 12 (50%) | 3 (4.5%) | 0 (0%) |
| | | Non-subject | 21 (17.9%) | 2 (8.3%) | 6 (9%) | 0 (0%) |
| | Inter. | Subject | 19 (16.2%) | 3 (12.5%) | 0 (0%) | 0 (0%) |
| | | Non-subject | 5 (4.3%) | 1 (4.2%) | 0 (0%) | 1 (9.1%) |
| Noun | Intra. | Subject | 1 (0.9%) | 1 (4.2%) | 0 (0%) | 2 (18.2%) |
| | | Non-subject | 2 (1.7%) | 0 (0%) | 4 (6%) | 0 (0%) |
| | Inter. | Subject | 17 (14.5%) | 3 (12.5%) | 5 (7.5%) | 1 (9.1%) |
| | | Non-subject | 6 (5.1%) | 2 (8.3%) | 7 (10.4%) | 1 (9.1%) |
| | Total | | 117 (100%) | 24 (100%) | 67 (100%) | 11 (100%) |

Table 11: Anaphora in subordinate clauses

As expected, in English zero anaphora was only used in coordinate clauses to select the subject of the previous clause. In Spanish, this was the case for around 40 per cent of ellipses, but, regardless of the clauses, more than 86 per cent of them were used to select a subject antecedent.

Participants mostly used subject pronouns (usually in a main clause) to retrieve subject antecedents, except for Spanish native speakers, who only used one subject pronoun in a subordinate clause to retrieve an intersentential non-subject antecedent, as shown in (9). The majority of anaphoric common and proper nouns were in main clauses, and they also recovered another subject. Spanish learners, however, used nominal anaphora mainly to retrieve non-subject antecedents.

- (9) Una vez con el estómago lleno, le contaron la tragedia a la **vecina**, la cual sin dudarle un momento los invitó a vivir con ella. Su esposo y dos hijos habían muerto en la guerra, así que **ella** también estaba sola. (sn3) ‘Once with a full stomach, they told the **neighbour** about the tragedy, who without doubting for a moment invited them to live with her. Her husband and two children had died in the war, so **she** was alone too.’

In summary, there were many similarities between Brazilian learners of English and Spanish learners of English in the production of anaphora. To start, around 45 per cent of the anaphors were in main clauses, which also meant that more than 50 per cent of the anaphors had an intersentential antecedent.

As expected, most anaphors in both languages (around 70%) continued the topic by retrieving the previous subject. More than half of the subject anaphors were pronouns in English and ellipsis in Spanish, and more than 70 per cent of them were used for topic continuity. Nouns and pronouns were used to shift the topic, and most cases (more than three fourths) did not have any pragmatic issues (i.e. ambiguity and redundancy). However, more than one fifth of the anaphors produced by Spanish learners were considered redundant. We noted that, in spite of the preferences for pronominal or zero anaphora, English and Spanish learners behaved in a similar way. Still, there were some differences between them and native speakers. English native speakers used more main clauses, while Spanish native speakers preferred to coordinate clauses. Compared to native speakers, it was also clear that learners could use more frequently zero anaphora instead of pronominal anaphora in English coordinate clauses and in Spanish main clauses to continue the topic. This was addressed during the online course and will be discussed in Section 4.3.

4.3. Differences between the instructional modes (synchronous, asynchronous, and control) on the learning of anaphora

To answer the third research question, we will discuss some differences in the use of anaphora by the experimental and control groups over time. We will first analyse the English groups and then the Spanish ones.

4.3.1. The English groups

As stated in previous sections, English learners were not ambiguous and were not much redundant in their texts. Still, they were instructed about how to omit the subject expression when possible and, as Figure 3 shows, the number of redundant anaphors decreased for both experimental groups.

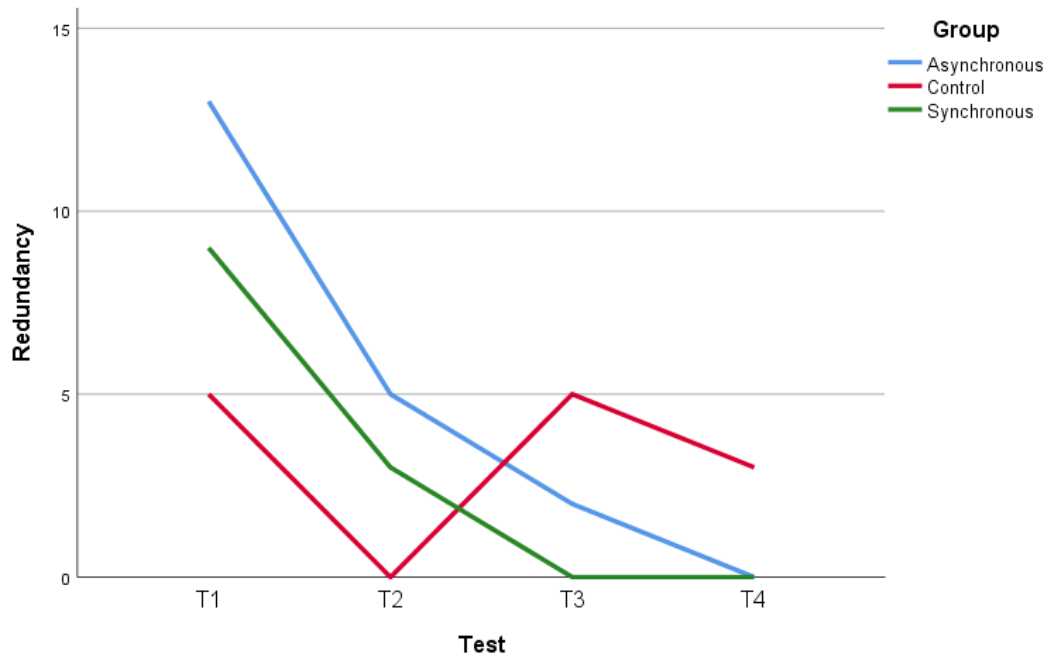


Figure 3: Redundancy in English

Although English learners were not so redundant in the first test compared to native speakers, they could still learn to use more zero anaphora in coordinate clauses with the same subject. They studied how to do it during the course and changed their anaphoric behaviour.

Figures 4 and 5 show that the asynchronous group started to use more zero anaphora in coordinate clauses, as well as less pronominal anaphora in main clauses to continue the topic, especially until the second post-test. This possibly happened because this group of learners chose to coordinate more clauses instead of separating them in different sentences.

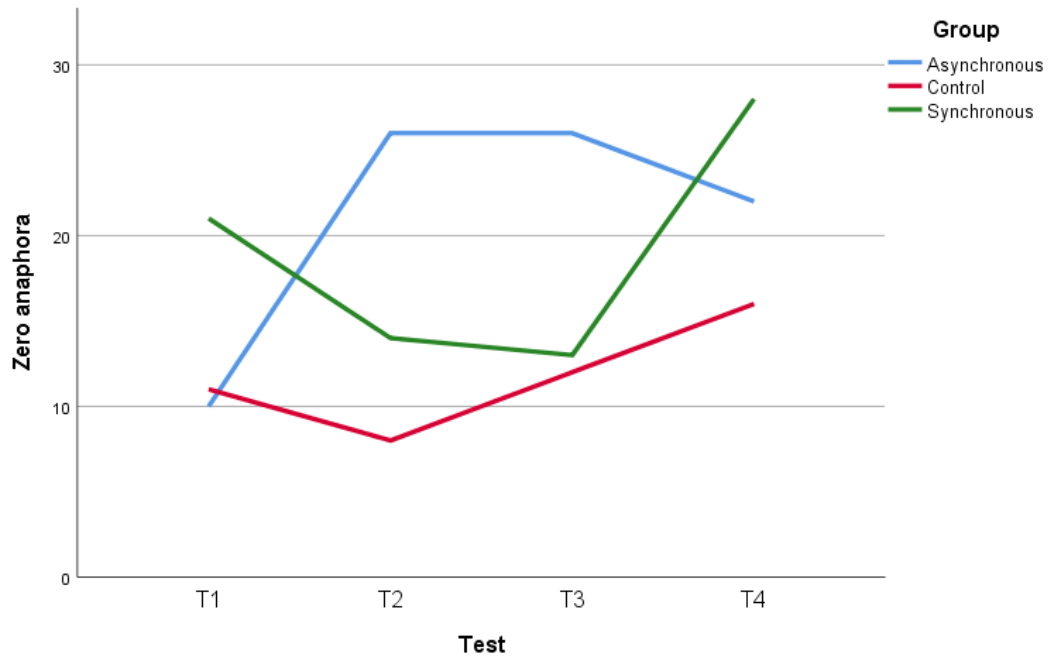


Figure 4: Zero anaphora in English coordinate clauses to retrieve the previous subject

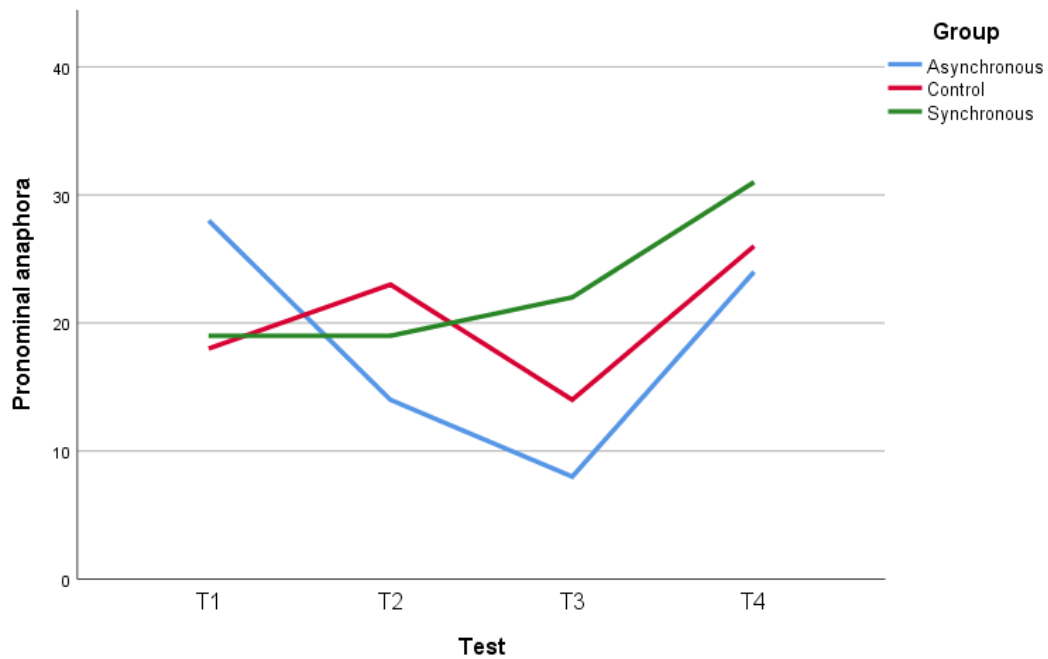


Figure 5: Pronominal anaphora in English main clauses to retrieve a previous subject

The data in Figure 6 also indicate a decrease in the use of pronominal anaphora by the synchronous group in coordinate clauses when there was topic continuity, but, as with the asynchronous group, the changes happened mainly until the third test. In the final test, one month after the course, the results were more similar to the pre-test.

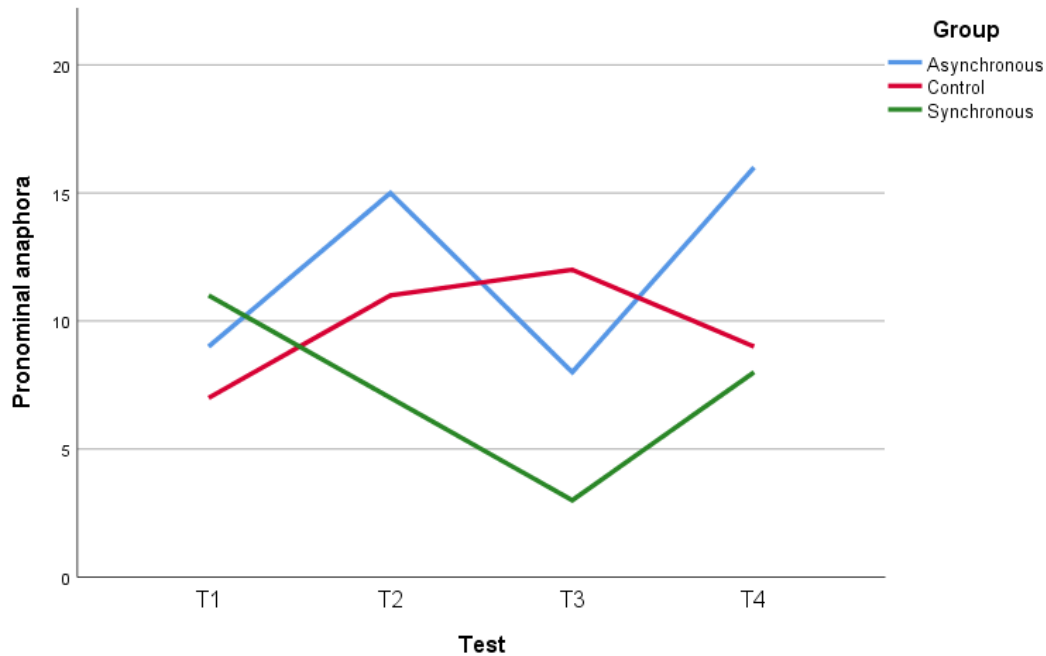


Figure 6: Pronominal anaphora in English coordinate clauses to retrieve the previous subject

In view of these findings, we can conclude that both experimental groups learned to be less redundant and to use less pronominal anaphora for topic continuity. However, only the asynchronous groups started to use more zero anaphora, and the changes were mainly until the second post-test.

4.3.2. The Spanish groups

In general, Spanish learners behaved very similarly to native speakers, probably because Portuguese and Spanish are both null-subject languages. As proposed by Lozano (2016) and confirmed in our study, learners were not ambiguous, but redundant in their texts. To solve this issue, during the course they studied how to use more zero anaphora instead of pronominal anaphora when continuing the topic and thus reduced the number of redundant anaphors, as can be seen in Figure 7.

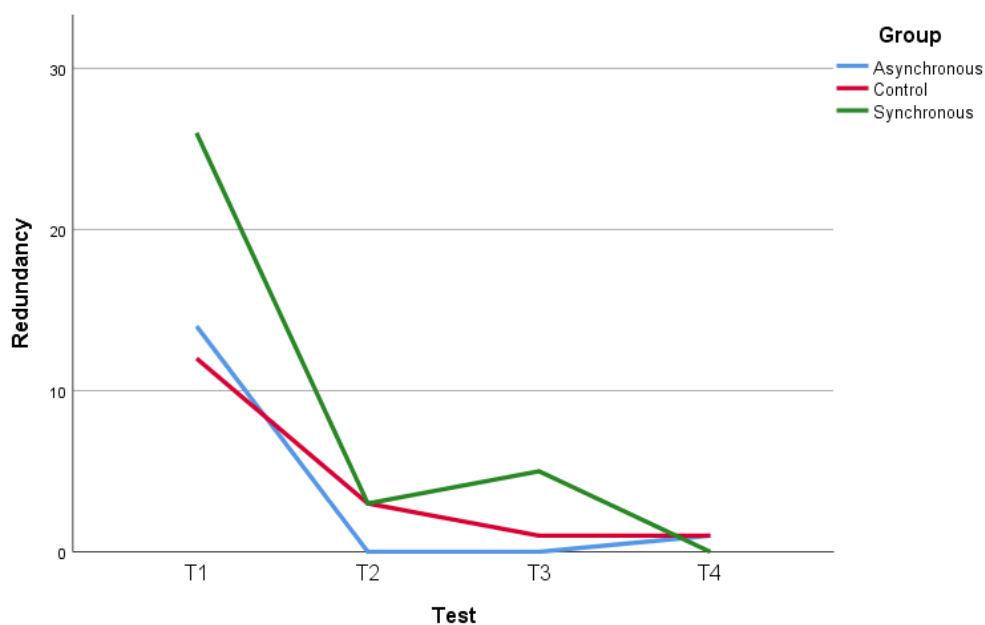


Figure 7: Redundancy in Spanish

Based especially on the results from main clauses, Spanish learners could also learn to use more zero instead of pronominal anaphora. They studied it during the course and, as Figures 9 and 9 present, the asynchronous group increased the use of zero anaphora until test 3, as well as continuously decreased the use of pronominal anaphora. Although the synchronous group also used fewer pronouns in test 2, the numbers increased in the following tests.

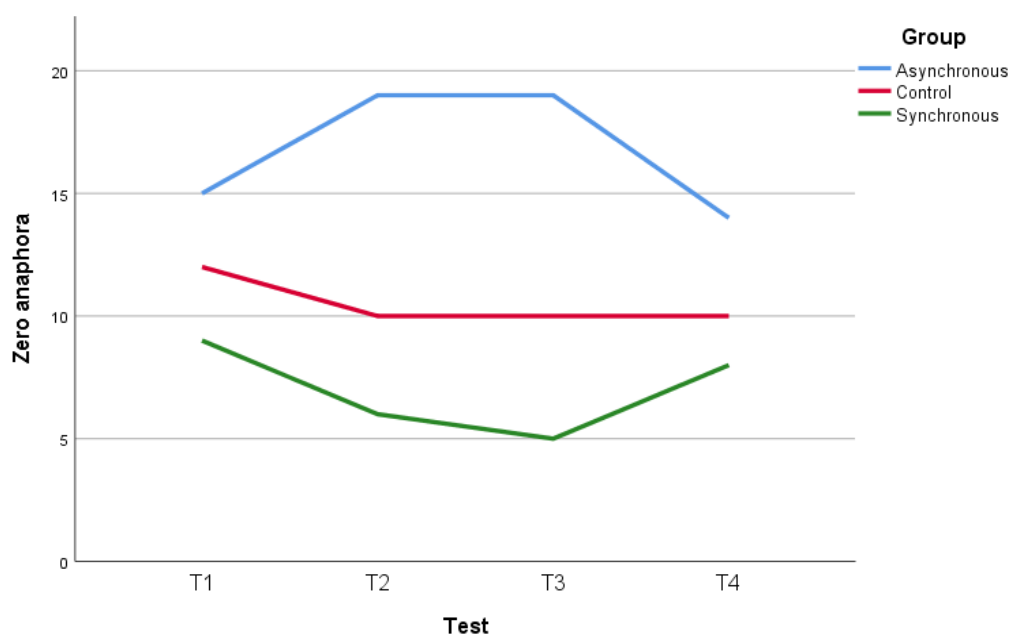


Figure 8: Zero anaphora in Spanish main clauses

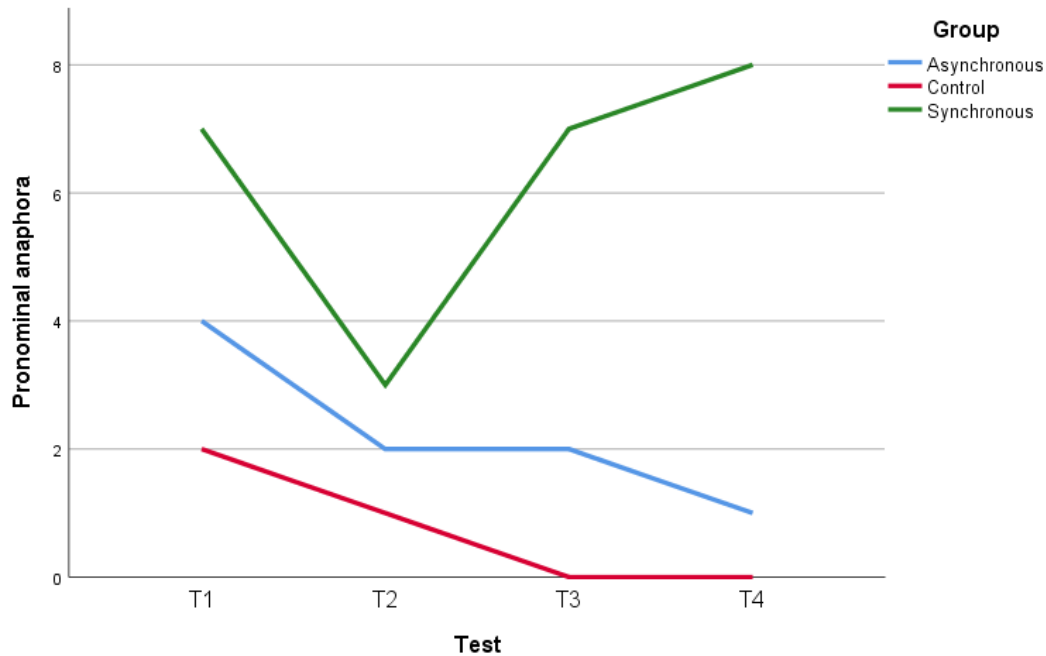


Figure 9: Pronominal anaphora in Spanish main clauses

In general, the asynchronous group performed better than the other groups. Regarding the increase of zero anaphora in Spanish main clauses and English coordinate clauses, the asynchronous group differed from the synchronous and control groups in the second and third tests.⁷ The synchronous and control groups did not present a significant difference between each other, and the three groups behaved similarly in the initial pre-test and the final post-test.

5. CONCLUSION

The present paper investigated the use of nominal, pronominal, and zero anaphora in two written corpora: BRANEN and BRANES. We designed an online course in two different instructional modes (synchronous and asynchronous) to investigate their impact on the learning of anaphora in English and Spanish over time. There were 45 participants (including control groups) who wrote narrative texts in four moments. Based on Lozano's (2016) annotation scheme of subject expressions, we annotated manually third-person

⁷ The Kruskal-Wallis test performed in SPSS (v. 26) only identified statistically relevant differences from the asynchronous group and the other groups on the second [$X^2(2) = 6.234$; $p = 0.044$] and third tests [$X^2(2) = 8.054$; $p = 0.018$]. The limited sample size does not allow for further elaboration.

human subjects of finite clauses and their antecedents in the texts using the *Recogito* annotation platform.

Since there was a small number of Spanish speakers (15 learners and three natives), we compared the use of zero and pronominal anaphora in BRANES and CEDEL2 (Lozano 2021b) before we started analysing and interpreting the results. Coincidentally, the percentages found were identical between the two groups of learners, and extremely similar between the groups of natives. Thus, we could answer RQ1 and consider our corpora representative.

After attesting the representativeness of the corpora, we analysed how Brazilian speakers processed anaphora in English and Spanish as foreign languages before the intervention (Tables 6 to 11) to answer the RQ2. We found similarities between learners and native speakers, which could be explained by the apprentices' intermediate-to-advanced level of proficiency in the language, but learners' distribution of anaphora in the types of clauses was much alike, regardless of the target language. It is possible that they have been more influenced by their L1 syntax than by the L2 when writing the sentences (as already suggested by Bruscato and Baptista 2021d, 2022a, 2022b regarding learner's reading strategies). Our study also confirmed Lozano's (2016) hypothesis, according to which learners are more redundant than ambiguous. In the pre-test, participants could have used more zero anaphora instead of pronominal anaphora in English coordinate clauses and in Spanish main clauses to continue the topic.

Finally, we investigated the effect of the instructional modes (synchronous and asynchronous) on the learning of the discursive mechanism (Figures 3 to 9) to answer RQ3. Although both experimental groups showed progress on the learning of anaphora, contrary to the control groups, the results revealed that the asynchronous instructional mode was more effective, probably because learners had more opportunities to read and write on the written forums than the synchronous groups on the oral discussions, but only until the third test.

In spite of the interesting remarks made above, the current study had some limitations that must be acknowledged. Firstly, the corpora contain only 180 short texts written by 45 learners with an intermediate-to-advanced level in the foreign language, a somewhat limited sample considering all possible target subjects. It would be interesting to compare the results here with data from more informants and with different levels of

proficiency. Besides, this was a short course, and the experimental groups had only two lessons on anaphora. In the future, the duration of the course could be extended, and it could include more testing moments. The effectiveness of the course over a longer period could also be investigated. Finally, our research focused only on third-person human subject anaphors, but non-human subjects or even verb complements could be annotated and analysed. To this end, the data retrieved from BRANEN and BRANES can be put to good use.

The major contribution of this paper is to show that Brazilian learners of Spanish and English use anaphora differently in relation to their instructional mode and to provide the scientific community with real, textual data for further investigation. To the best of our knowledge, a distant learning mode-specific approach to anaphora learning like that had not been described yet. In the future, besides pursuing some of the lines of research already sketched above, we also plan to investigate the impact of synchronous and asynchronous learning to the understanding and the production of anaphora in spoken texts.

REFERENCES

- Andreu-Andrés, M^a Ángeles, Aurora Astor, María Boquera, Penny MacDonald, Begoña Montero and Carmen Pérez. 2010. Analysing EFL learner output in the MiLC project: An error it's*, but which tag. In Mari Campoy-Cubillo, Begoña Bellés-Fortuño and María Lluisa Gea-Valor eds. *Corpus-based Approaches to English Language Teaching*. London: Continuum, 167–179.
- Baptista, Jorge, Simone Pereira and Nuno Mamede. 2016. ZAC: Zero Anaphora Corpus. In João Silva, Ricardo Ribeiro, Paulo Quaresma, André Adami and António Branco eds. *Proceedings of the International Conference on Computational Processing of the Portuguese Language 2016*. Tomar: Springer International Publishing, 38–45.
- Bruscatto, Amanda Maraschin and Jorge Baptista. 2021a. BRANEN and BRANES Corpora. In *Proceedings of the 9th European Conference on Language Learning*, London: IAFOR, 1–13. <https://doi.org/10.22492/issn.2188-112X.2021.3>
- Bruscatto, Amanda Maraschin and Jorge Baptista. 2021b. Designing an online course to teach anaphora in foreign languages. *International Journal of Second and Foreign Language Education* 1/1: 1–9.
- Bruscatto, Amanda Maraschin and Jorge Baptista. 2021c. Synchronous and asynchronous distance learning of anaphora in foreign languages. *Texto Livre* 14/1: 1–16.
- Bruscatto, Amanda Maraschin and Jorge Baptista. 2021d. Resolução de ambiguidade anafórica em português, inglês e espanhol (estudo-piloto). *Alfa* 65/1: 1–22.
- Bruscatto, Amanda Maraschin and Jorge Baptista. 2022a. Resolução de ambiguidade anafórica em português, inglês e espanhol: Comparação de dados obtidos com falantes de PE e PB. *Domínios de Linguagem* 16/1: 1–29.
- Bruscatto, Amanda Maraschin and Jorge Baptista. 2022b (in press). The resolution of

- ambiguous anaphora in English, Spanish, and Portuguese. *Delta*.
<https://doi.org/10.1590/1678-460x202151581>
- Chomsky, Noam. 1981. *Lectures on Government and Binding: The Pisa Lectures*. Dordrecht: Foris.
- Ellis, Rod. 2008. *The Study of Second Language Acquisition*. Oxford: Oxford University Press.
- Ghaddar, Abbas and Philippe Langlais. 2016. Wikicoref: An English coreference-annotated corpus of Wikipedia articles. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Sara Goggi, Marko Grobelnik, Bente Maegaard, Joseph Mariani, H el ene Mazo, Asunci on Moreno, Jan Odijk and Stelios Piperidis eds. *Proceedings of the Language Resources and Evaluation Conference, LREC 2016*. Portoro : European Language Resources Association, 136–142.
- Granger, Sylviane. 2003. The International Corpus of Learner English: A new resource for foreign language learning and teaching and second language acquisition research. *Tesol Quarterly* 37/3: 538–546.
- Grice, Herbert Paul. 1975. Logic and conversation. In Peter Cole and Jerry Morgan eds. *Syntax and Semantics, Vol. 3: Speech Acts*. New York: Academic Press, 41–58.
- Hendriks, Henri ette. 2003. Using nouns for reference maintenance: A seeming contradiction in L2 discourse. In Anna Ramat ed. *Typology and Second Language Acquisition*. Berlin: Mouton de Gruyter, 291–326.
- IBM Corporation. 2020. *IBM SPSS Statistics for Windows, version 26*. Armonk: IBM Corporation.
- Kilgarriff, Adam, V it Baisa, Jan Bu ta, Milo  Jakub icek, Vojt ech Kov ar, Jan Michelfeit, Pavel Rychl y and V it Suchomel. 2014. The Sketch Engine: Ten years on. *Lexicography* 1/1: 7–36.
- Li, Liu. 2014. Computer-assisted vs. classroom instruction on developing reference tracking skills in L2 Chinese. In Shuai Li and Peter Swanson eds. *Engaging Language Learners through Technology Integration: Theory, Applications, and Outcomes*. Hershey: Information Science Reference, 72–96.
- Liu, Rong. 2010. *The Acquisition and Online Processing of Anaphora by Chinese-English Bilinguals: A Computer Assisted Study*. Arizona: University of Arizona dissertation.
- Lozano, Crist bal. 2016. Pragmatic principles in anaphora resolution at the syntax-discourse interface: Advanced English learners of Spanish in the CEDEL2 corpus. In Margarita Alonso-Ramos ed. *Spanish Learner Corpus Research: Current Trends and Future Perspectives*. Amsterdam: John Benjamins, 235–265.
- Lozano, Crist bal. 2021a. Anaphora resolution in Second Language Acquisition. In Mark Aronoff ed. *Oxford Bibliographies in Linguistics*. Oxford: Oxford University Press.
<https://www.doi.org/10.1093/obo/9780199772810-0268>
- Lozano, Crist bal. 2021b. CEDEL2: Design, compilation and web interface of an online corpus for L2 Spanish acquisition research. *Second Language Research*, first online view. <https://doi.org/10.1177/02676583211050522>
- Lozano, Crist bal, Ana D az-Negrillo and Marcus Callies. 2020. Designing and compiling a learner corpus of written and spoken narratives: COREFL. In Christiane Bongartz and Jacopo Torregrossa eds. *What’s in a Narrative? Variation in Story-telling at the Interface between Language and Literacy*. Bern: Peter Lang, 21–46.
- Marques, Jo o. 2013. *Anaphora Resolution in Portuguese: A Hybrid Approach*. Lisbon: Universidade de Lisboa dissertation.
- Miltsakaki, Eleni. 2002. Toward an aposynthesis of topic continuity and intrasentential

- anaphora. *Computational Linguistics* 28/3: 319–355.
- Pardo, Thiago Alexandre Salgueiro, Jorge Baptista, Magali Sanches Duran, Maria das Graças Volpe Nunes, Fernando Antônio Asevedo Nóbrega, Sandra Maria Aluísio, Ariani di Felippo, Eloize Rossi Marques Seno, Raphael Rocha da Silva, Rafael Torres Anchiêta, Henrico Bertini Brum, Márcio de Souza Dias, Rafael de Sousa Oliveira Martins, Erick Galani Maziero, Jackson Wilke da Cruz Souza and Francielle Alves Vargas. 2017. The coreference annotation of the CSTNews corpus. In Raquel Martínez, Julio Gonzalo, Paolo Rosso, Soto Montalvo and Jorge Carrillo-de-Albornoz eds. *Proceedings of the Second Workshop on Evaluation of Human Language Technologies for Iberian Languages, vol. 1881*. Murcia: IberEval, 102–112.
- Poesio, Massimo and Ron Artstein. 2008. Anaphoric annotation in the ARRAU corpus. In Nicoletta Calzolari, Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, Daniel Tapias eds. *Proceedings of the Sixth International Conference on Language Resources and Evaluation*. Marrakech: European Language Resources Association (ELRA), 1170–1174.
- Pradhan, Sameer, Eduard Hovy, Mitch Marcus, Martha Palmer, Lance Ramshaw and Ralph Weischedel. 2007. Ontonotes: A unified relational semantic representation. In *Proceedings of the International Conference on Semantic Computing*. Irvine, California: IEEE Computer Society, 517–526.
- Quesada, Teresa and Cristóbal Lozano. 2020. Which factors determine the choice of referential expressions in L2 English discourse? A multifactorial study from the COREFL corpus. *Studies in Second Language Acquisition* 42/5: 959–986.
- Rizzi, Luigi. 1982. *Issues in Italian Syntax*. Dordrecht: Foris.
- Siemens, George, Dragan Gašević and Shane Dawson. 2015. *Preparing for the Digital University: A Review of the History and Current State of Distance, blended, and online learning*. Arlington: Link Research Lab.
- Tagnin, Stella. 2006. A multilingual learner corpus in Brazil. In Andrew Wilson, Dawn Archer and Paul Rayson eds. *Corpus Linguistics Around the World*. Brill Rodopi, 195–202.

Corresponding author

Amanda Maraschin Bruscato
 Gambelas Campus
 8005-139
 Faro Portugal
 e-mail: a66230@ualg.pt

received: November 2021
 accepted: June 2022
 published online: July 2022