

The interpretation of reflexive pronouns in VP-ellipsis by L2 English learners with different proficiency levels

Francisco Gallardo-del-Puerto, Evelyn Gandón-Chapela
Universidad de Cantabria

English reflexive anaphora in cases of VP-ellipsis may allow for strict and sloppy readings. A few L2 studies (Epoge, 2012; Park, 2016; Ying, 2005) have focused on determining the role that L2 proficiency may exert on learners' choices in bare, referential, and non-referential contexts. This paper provides data from 104 Spanish learners of English (A2, B1, and B2 levels) and 32 native speakers of English. Results showed that participants tended to interpret reflexives sloppily in bare and non-referential contexts, whereas strict readings prevailed in referential ones. There existed significant differences in the interpretation of learners versus native speakers, whilst the differences among the three learner groups were not so marked. However, the least proficient group differed most from native speakers. Findings partially confirm previous research and discrepancies may be tentatively ascribed to extraneous variables (e.g., the learners' L1, the range of the proficiency levels, or the characteristics of the control groups).

Keywords: VP-ellipsis, reflexive pronouns, strict reading, sloppy reading, proficiency level, L2 English

1. Introduction

Over the past five decades, the syntactic phenomenon of ellipsis, i.e., the omission of subcategorised linguistic elements thanks to the presence of linguistic or extralinguistic antecedents, has been subject of fruitful research both from a theoretical and an applied perspective. Thus, it is no surprise that studies on ellipsis have proliferated not only in theoretical frameworks such as Generative Grammar or Relevance Theory, but also in more applied fields such as First and Second Language Acquisition (SLA) or psycholinguistics. In the case of the present paper, its main aim is to provide an analysis

of the interplay between target language proficiency and the interpretation of reflexive anaphora in cases of Verb Phrase ellipsis (VPE) in the field of SLA. It should be noted that VPE “is relatively limited in its distribution across world’s languages compared to the more widespread sluicing” (Aelbrecht, 2009, p. 15) and it is not licensed in languages such as Spanish, Dutch, French or Italian. To be more precise, we will analyse the interaction between English learners’ proficiency levels and their reading of ambiguous sentences such as “Richard painted himself and James did too,” given that it could be the case that “James painted Richard” (strict reading) or that “James painted himself” (sloppy reading). Since, to our knowledge, this interplay has only been investigated with L1 Chinese (Ying, 2005), L1 Korean (Park, 2016) and Cameroon English speakers (Epoge, 2012), we explore the readings of reflexive pronouns in cases of VPE made by three different proficiency level groups (A2, B1, and B2) of L1 Spanish learners of English. Studying this phenomenon in Spanish is particularly relevant as there is a dearth of L2 English studies on this issue with learners whose L1 is a Romance language. This is a population which is worth investigating considering that Spanish is typologically closer to English than Chinese and Korean are.

In this paper, section 2 summarises the state of the art. In section 3, we will present the research questions. The methodology used to obtain the data will be described in section 4. Section 5 presents the results, which are subsequently discussed in section 6. Section 7 provides a summary of the main findings, concluding remarks, and issues for further research.

2. State of the art

In this section, we will first provide a general overview of the features exhibited by cases of VPE with reflexive anaphora in English and the interpretation ambiguities that may arise, as mentioned in the relevant literature. Following, we will offer a brief account of the L1 and L2 studies that have focused on this type of constructions. Lastly, we will refer to the impact that target language proficiency may exert on the readings that reflexive pronouns may receive in cases of VPE.

2.1 Reflexive anaphora in cases of VPE

As evinced in the literature, over the past decades VPE has been the type of ellipsis that has received most attention from researchers working in the Generative Grammar framework (see, for instance, Aelbrecht, 2010; Hankamer & Sag, 1976; Johnson, 2001; Williams, 1977; to name but a few). In particular, the interpretations of English reflexive pronouns in cases of VPE have been subject of hot theoretical linguistics debates (e.g., Dalrymple, 2005; Dalrymple et al., 1991; Fiengo & May, 1994; Hestvik, 1995; Kitagawa, 1991; McKillen, 2016; Sag, 1976) since reflexive anaphora in VPE can lead to two potential readings, as instantiated in (1):

- (1) Mary_i blamed herself_i and Heather_j did too.
- a. Heather_j blamed Heather_j (sloppy reading)
 - b. Heather_j blamed Mary_i (strict reading)

As can be observed in (1), the reflexive pronoun can be interpreted either sloppily (“Heather blamed Heather”) or strictly (“Heather blamed Mary”). In the literature, these two different readings have been traditionally explained by Principle A and Principle B of the Binding Theory (Chomsky, 1981). On the one hand, Principle A is responsible for the sloppy interpretation because it argues that an anaphor must be bound locally. Since the reflexive anaphor needs to be bound within its local domain, it will refer back to the subject of the second conjunct of the VPE construction, i.e., “Heather”, rendering the sloppy meaning of the ellipsis site as “Heather blamed Heather”. On the contrary, Principle B dictates that a (non-anaphoric) pronominal (expression) must be free within its local domain, allowing one to interpret the example in (1) strictly: “Heather blamed Mary”. This means that the anaphor in the second conjunct of the VPE construction would refer back to the subject of the first conjunct (“Mary”), which acts as the antecedent of the ellipsis site.

In the literature on the topic there seems to be no controversy concerning the sloppy reading that reflexives may receive in cases of VPE if one assumes that they function obligatorily as bound variables (Hestvik, 1995, p. 211). However, as put forward by McKillen (2016, p. 14), the fact that “reflexives can give rise to a strict reading [...] is unexpected in a binding theory that necessarily treats reflexives as locally bound variables”. Therefore, the centre of the debate has revolved around the

status of the strict interpretation. While for some authors only the sloppy reading was possible (e.g., Williams, 1977), for others, strict readings were available only in cases of non-locally bound reflexives (e.g., Bouchard, 1984). Still others argued that reflexives could be either strict or sloppy (e.g., Sag, 1976). As will be shown in what follows, this issue has been approached from two different perspectives: a semantic and a syntactic one.¹ Dalrymple et al. (1991) represent the former and claim that a strict interpretation may be available depending on the semantic property of individual verbs. For example, verbs such as “defend” and “lock” are regarded as different in the sense that whereas the former allows a strict interpretation, the latter does not (Ying, 2005, p. 552):

- (2) Bill defended himself against the accusation, and John did, too.
- (3) John locked himself in the bathroom when bad news arrived, but Bill would never do so.

In Dalrymple et al.’s (1991) approach to the issue, the verb “defend” in (2) may receive both a sloppy and a strict interpretation given that this type of verb does not need its subject and object to be co-referential. On the contrary, the verb “lock” in (3) would not license a strict reading because it requires its subject and object to be co-referential. This semantic approach was challenged by Hestvik (1995), who claimed that the availability of strict readings does not depend on the semantic characteristics of lexical items, but on their syntactic structure. More precisely, he contended that while subordination of the elided clause with respect to the antecedent clause facilitates a strict reading, coordination disfavours it. This is illustrated in examples (4) and (5), in which, as will be shown, both the verb “defend” and “lock” allow for a strict interpretation of the ellipsis sites:

- (4) a. John defended himself better than Bill did.
b. John defended himself, and Bill did too.
- (5) John locked himself in the bathroom before Bill could.

¹ For thorough overviews on the availability of strict and sloppy readings for reflexives in cases of VPE, see Hestvik (1995) and McKillen (2016).

As put forward by Hestvik (1995, p. 233), Dalrymple et al.'s prediction would be that the verb "defend" "should always be able to get a strict reading, no matter what the structure is". However, he claims that the strict interpretation ("Bill defended John") is better in (4a) than in (4b). Concerning the verb "lock", according to Dalrymple et al. (1991), it should not allow a strict reading irrespective of what the structure is. Hestvik's (1995, p. 233) example in (5) shows that this is not confirmed, which leads him to claim that since a strict reading is possible, syntactic "structure does play a role". However, he admits that "[a]lthough it is true that some verbs favor a sloppy interpretation, once this lexical semantic effect is controlled for, structure determines strict/sloppy readings independently of verb choice" (Hestvik, 1995, p. 233). More recently, McKillen (2016, p. 17) has shown that coordinated ellipsis with a strict reading does sound natural if the right context is provided:

- (6) *Context: John and Bill are very good friends, and would do anything to help the other out. When John was wrongfully accused of stealing some office supplies...
...he defended himself and Bill did too= John defended John and Bill defended John (strict reading)*

Consequently, an approach to the issue based on the semantic properties of the different verbs was shown to fail to account for the data. This issue was also tackled by Kitagawa (1991), who proposed a reconstruction of reflexive pronouns at LF.² She defended that a feature [+anaphor] could be removed when copying the VP that acts as the antecedent into the ellipsis site. This would allow the reflexive that acts as the antecedent to be reconstructed as a pronoun. In (7) the LF of (7a) would be (7b) under her analysis:

- (7) a. John_i likes himself_i, and Bill does too.
b. John_i likes [_{+a}] himself_i, and Bill likes [_{-a}] him_i too.

This idea was further developed by Fiengo and May (1994), who proposed a syntactic account of the phenomenon, which they coined "vehicle change". The main idea behind their proposal was that it is possible for a reflexive to change to a pronoun when it is copied from the antecedent to the ellipsis site. Therefore, the strict reading is possible thanks to the reconstruction of the reflexive as a pronoun (vehicle change), which,

² In Generative Grammar, LF stands for 'Logical Form', a level of mental representation of a linguistic expression which is derived from surface structure.

following Principle B of the Binding Theory, cannot be locally bound. Hence, this account provides an explanation as to why the reflexive pronoun “himself”, present in the antecedent VP in (7), may be reconstructed as the pronoun “him” in the ellipsis site, therefore co-referring with the nonlocal subject “John”.

Given that both syntactic and semantic approaches to the type of identity required between the antecedent and the ellipsis site failed to account for the whole range of data, there emerged other proposals, presented, for instance, in Kehler (2000). This author tried to account for the data by resorting to discourse structure and not to syntactic structure. He argued that (un)acceptable examples of identity mismatch could be classified by the type of coherence relation established between the antecedent and the ellipsis site, namely, Resemblance and Cause-Effect relations. More precisely, he contended that when a Resemblance coherence relation holds, where the similarity between the parallel events described in the two clauses is emphasised, identity mismatch is not allowed. This is due to the fact that the two events described in the antecedent and the ellipsis site are interpreted as independent and thus receive a parallel or symmetric reading (see example (8) below). In contrast, when the coherence relation established between the two clauses is that of Cause-Effect, what is highlighted is the causal relation between the events reported in the two clauses, as in example (9) below. In this case, the first event is interpreted as the cause of the second event, that is, there is an asymmetric reading. This leads Kehler (2000) to argue that syntactic identity is only required in cases of Resemblance coherence, but not in cases of Cause-Effect coherence. Therefore, Kehler’s theory (2000) would predict grammatical violations of the binding theory in cases where the Cause-Effect discourse coherence relation holds (as in (9)), but not in those examples where the Resemblance coherence type holds (as in (8)):

(8) *John’s_i lawyer defended him_i, and he_i did too. [defend him_i] [Kehler (2000, pp. 550-54)]

(9) John’s_i lawyer defended him_i because he_i wouldn’t. [defend him_i]

In (9) condition B of the binding theory is violated, since it allows the coreference between the noun “John” and the pronoun “him” in the target of ellipsis, where “him” is interpreted as “himself.” However, according to Kehler (2000, p. 554), coreference between these two items is not allowed in (8).

More recently, in her dissertation, McKillen (2016) questioned the Standard Binding Theory (Chomsky, 1981) assumption that reflexives can be interpreted only as bound variables (as set by Principle A) by investigating the possible readings of reflexives in cases of VPE and focus constructions. This author contends that the fact that reflexives may receive a strict and a sloppy interpretation arises the question of whether Principle A's requirement that reflexives be interpreted only as bound variables is correct (McKillen, 2016, p. 15). What is more, she also refers to a fact that had already been noted by Ross (1967, p. 348), which is that pronouns in ellipsis constructions are ambiguous, not only reflexives:

(10) a. John₁ defended his₁ brother and Bill did too.

b. John defended John's brother and Bill defended John's brother (strict)

(11) a. John₁ defended his₁ brother and Bill did too.

b. John defended John's brother and Bill defended Bill's brother (sloppy)

As can be observed, both strict and sloppy readings are available in the previous examples, which do not involve reflexive pronouns. In the literature, the availability of these two different readings has been claimed to derive from the pronoun's ability to be interpreted either referentially (that is, as a free variable), giving rise to the strict reading, or as a bound variable, making the sloppy reading possible (McKillen, 2016, p. 16). Since, as shown in some of the previous examples, both sloppy and strict readings are available for reflexives in elliptical constructions, McKillen (2016) argues that this constitutes evidence for both referentially interpreted and bound-variable reflexives, as happens with ambiguities that arise with non-reflexive pronouns. Therefore, even though previous accounts such as Hestvik (1995) had opted for deriving strict readings of reflexives while defending the bound-variable-only assumption, McKillen (2016) reformulated Condition A of the Binding Theory to incorporate the strict interpretation.

In summary, theoretical linguistics over the past decades has tried to offer an analysis that accounts for the availability of both strict and sloppy interpretations in cases of VPE with reflexive pronouns, but, as Ying argues (2005, p. 553), "neither account tells us how readers would interpret this strict-sloppy ambiguity". Thus, the interpretation that these types of constructions may receive by native and non-native speakers of English has been object of the present paper.

2.2 L1 English studies on reflexive anaphora in VPE

Both acquisition and processing studies on VPE have mainly investigated parallelism effects (VPE vs. VP anaphora³ and VPE vs. Gapping⁴) and pronoun interpretation (Hwang, 2020). Within the field of L1 acquisition, most investigations have been carried out from an experimental perspective and focused on children's early production and comprehension of VPE in English (Postman et al., 1997; Thornton & Wexler, 1999).

Crucially, for the present paper, there have also been some studies that have tested the interpretation of reflexive anaphora (either as strict or sloppy) in cases of VPE by adult native speakers of English thanks to judgement tasks (Frazier & Clifton, 2006; Gandón Chapela & Gallardo del Puerto, 2019; Kim & Runner, 2009; Ong & Brasoveanu, 2014; Park, 2016; Ying, 2005). Frazier and Clifton (2006) tested Kehler's (2000) discourse coherence theory empirically in various processing experiments with a two-fold purpose. Firstly, they intended to figure out whether Principle A of the Binding theory must be obeyed in cases of ellipsis involving Resemblance relations (as in (12a)) but not in those involving Cause-Effect relations (as in (12b)). Secondly, they also tested the effect imposed by syntactic parallelism between the antecedent clause and the one containing the ellipsis site on the preference for either a strict or a sloppy interpretation of pronominals or reflexives (see example (13)).

(12) a. Doug blamed himself for the band's collapse just like everyone else did.

b. Doug blamed himself for the band's collapse because everyone else did.

(Frazier & Clifton 2006, p. 326)

(13) a. John saw a snake near him and Bill did too.

b. John saw a snake near his backpack and Bill did too.

(Frazier & Clifton 2006, p. 329)

The results of their experiments showed that strict readings were as likely to take place in Cause-Effect sentences as in Resemblance sentences. Also, they proved that if the

³ See Matsuo and Duffield (2001), Duffield and Matsuo (2009).

⁴ See Hwang (2020).

coherence relation was held constant and parallelism was increased by means of the “and...did too” construction, the number of sloppy interpretations for pronouns, possessive pronouns and reflexives increased. In conclusion, this study has shown that Kehler’s (2000) theory does not have an empirical basis. The lack of a syntactically appropriate antecedent affects the degree of acceptability of all sentences containing VPE, not just those involving Resemblance relations. Moreover, it has been shown that parallelism is favoured both in Cause-Effect and Resemblance relations. The solution that Frazier and Clifton (2006) offer in order to avoid the overgeneration problem (because semantic accounts of ellipsis overgenerate and syntactic ones undergenerate acceptable examples of ellipsis) is to propose a theory of processing which includes a syntactic account of ellipsis where the processor “patches up an antecedent at LF if it is not the required form” (Frazier & Clifton 2006, p. 15).

Kim and Runner (2009), on their part, put Hestvik’s (1995) and Kehler’s (2000) theories to the test in a two factorial design, as shown in sentences (14a-d) below:

(14) a. *Within-sentence, Resemblance:*

Jane voted for herself in the election, and Ann did too.

b. *Within-sentence, Cause-Effect:*

Jane voted for herself in the election, so Ann did too.

c. *Cross-sentence, Resemblance:*

Jane voted for herself in the election. Ann did too.

d. *Cross-sentence, Cause-Effect:*

Jane voted for herself in the election. So Ann did too.

The focus of their work was on the effect that discourse connectives exert on strict vs. sloppy readings of reflexives. Results indicated that the proportion of strict identity responses was surprisingly high in all conditions (Kim & Runner 2009, p. 279).

However, as predicted by the syntactic account, there were more strict interpretations within sentences than across sentences. In addition, as predicted by Kehler’s (2000) discourse theory, the rate of strict interpretations in sentences containing a Cause-Effect coherence relation was greater than in those containing a Resemblance coherence relation.

Ong and Brasoveanu (2014) extended Kim and Runner’s (2009) work by including three more conditions in order to compare the syntax vs. discourse-driven

account empirically. First, their study expanded the type of connectives used (“and,” “if-then”, and “so”) while keeping the syntactic configuration constant. Second, they also tested the role of negation in biasing a strict or a sloppy reading in combination with a set of discourse connectives (“and,” “but,” “(and) therefore,” “(but) nevertheless”). Finally, these authors tackled the effect of verb semantics in triggering the Causality Effect by distinguishing between Implicit Causality 1 verbs (where the subject initiates an action that leads to an emotional state in the object, as in “John disappointed Bill because he stole the book”) and Implicit Causality 2 verbs (where the object initiates an action that leads to an emotional state in the subject, as in “John scolded Bill because he stole the book”). The results of their experiments indicate that Kehler’s (2000) discourse-driven theory offers a more adequate explanation of the data than the syntactic account in determining the probability of strict vs. sloppy readings. Crucially, they also discovered that the meaning of the elided verb exerted an influence on the availability of both types of interpretations, strict readings being more likely to happen in object-oriented implicit causality verbs (type 1) than in subject-oriented ones (type 2). Finally, these authors also discovered that the position of sentential negation was a biasing factor with respect to the availability of strict vs. sloppy readings (early negation: “John didn’t blame himself and/but/nevertheless/therefore Bill did”; late negation: “John blamed himself and/but/nevertheless/therefore Bill didn’t”), with the former reading being significantly higher in cases of early sentential negation together with connectives “but” and “nevertheless” (although this was the case across all connectives).

Importantly, for the present paper, there have also been some studies which have tested the interpretation of reflexive anaphora by native speakers of English (control group) and Chinese (Ying, 2005), Korean (Park, 2016), and Spanish (Gandón Chapela & Gallardo del Puerto, 2019) learners of English using judgement tasks. In these studies, English native speakers were found to favour sloppy readings in bare (see example (15a)) and non-referential contexts (as in (15b)), while strict readings prevailed in referential contexts (see (15c)).

(15) a. Mary blamed herself and Heather did too.

b. Mary blamed herself and Heather did too. Heather has two brothers and a sister.

c. Mary blamed herself and Heather did too. Heather thinks Mary is a disaster.

As can be gathered, the ellipsis sites of examples (15a) and (15b) may be interpreted either strictly (“Heather blamed Mary”) or sloppily (“Heather blamed Heather”). However, in the case of example (15c) the contextual information available causes a bias in the interpretation of the ellipsis site towards a strict reading (“Heather blamed Mary”).

2.3 L2 English studies on reflexive anaphora in VPE⁵

As in the L1 literature, L2 acquisition studies have focused on testing L2 learners’ VPE comprehension (Kim, 2015), as well as on exploring the impact of various L1s on parallelism effects in cases of VPE (Al-Thubaiti, 2018), VPE vs. VP anaphora (Duffield & Matsuo, 2009) and VPE vs. Gapping (Hwang, 2020) in L2 English. These studies are not conclusive as regards whether it is transfer or Universal Grammar (UG) that accounts for L2 learners’ behaviour concerning the features of the aforementioned constructions, whose implicit knowledge cannot be acquired on the basis of L2 input only. Only Hwang (2020) seems to support the idea that UG is responsible for the results found in her study. She purports that her L1 Korean learners of English cannot rely on either English or Korean input alone in order to acquire implicit knowledge of the impossibility of both Gapping in adjunct clauses and the object reading for VPE in English.⁶

⁵ For more detailed information on the interplay between ellipsis and acquisition see Roeper (2018).

⁶ Hwang (2020, p. vii) explains that there are some contrasts between Gapping and VPE, which are two similar elliptical constructions in English. She mentions, for instance, that while VPE is possible both in conjunct clauses (*Sara made pizza and Kelly did too*) and in adjunct clauses (*Sara made pizza because Kelly did*), Gapping is grammatical only in conjunct clauses (*Sara made pizza {and Kelly pasta/*because Kelly pasta}*). Another contrast is that whereas Gapping (*Mom hugged the boy at home and Dad in the park*) allows the noun phrase that follows the conjunction to be read as either

Additionally, some L2 studies have tackled the ambiguity of interpretations in cases of reflexive anaphora,⁷ as illustrated in the following examples:

(16) John thought that Bill praised himself (Yip & Tang 1998, p. 175)

(17) Pinocchio is telling Donald Duck to point at himself/him (Lee & Schachter, 1997, p. 362)

(18) Simon says Jack should point to himself (Al Kafri, 2013, p. 118)

Scholarly discussions on this issue have mainly been broached from the perspective of the Generative framework in an attempt to elucidate whether, following the principles of the Binding Theory, parameter resetting is possible and thus UG is accessed by adult L2 learners. Research findings have been contradictory with supportive evidence in favour of both the Full Access Hypothesis (Bennett, 1994; Yip & Tang, 1998) and the Indirect Access Hypothesis (Al Kafri, 2013), but also in favour of an intermediate binding which is neither L1-like nor L2-like but still UG-constrained (Hirakawa, 1990; MacLaughlin, 1998). Further, the age factor has also been proven controversial, with authors supporting (Lee & Schachter, 1997) and rejecting (Al Kafri, 2013) the critical period hypothesis.

With respect to the interplay between VPE and reflexive anaphora, L2 research has centred its attention on English learners with various L1 backgrounds, namely, Chinese (Ying, 2005), Cameroon English (Epoge, 2012), Korean (Park, 2016), and, more recently, Spanish (Gandón Chapela & Gallardo del Puerto, 2019). The aim of these works has been to explore L2 learners' interpretations of reflexive pronouns (either as strict or sloppy) in cases of VPE taking into account three different contexts: bare, referential, and non-referential (see the examples illustrated in (15)). The results of Gandón Chapela and Gallardo del Puerto's (2019) study, which investigated forty-four L1 Spanish university learners of English, showed a fluctuation between sloppy and strict readings of English reflexives in the three VPE contexts. These learners, similarly to the control group of English native speakers, favoured sloppy readings in bare and non-referential contexts, whereas strict readings prevailed in referential contexts.

the subject (i.e., 'hugger') or object (i.e., 'huggee') of the gapped verb, VPE (*Mom hugged the boy at home and Dad did too*) licenses only a subject reading.

⁷ See Lozano (2021).

However, learners' rate of strict readings in referential contexts was significantly lower than that of native speakers. Moreover, learners' preference for sloppy readings decreased in non-referential contexts if compared with bare contexts. On the contrary, the presence of a non-referential context led native speakers to reinforce the sloppy reading. These authors interpreted these findings as showing that the L2 learners' interpretation of VPE is less straightforward since the added contextual information involving both linguistic and pragmatic processing acts as a distractor, triggering a further cognitive load.

There have also been some L2 studies that have analysed the effect of target language proficiency (defined as "overall L2 competence" (Xiao, 2015, p. 558)) on the acquisition of several linguistic phenomena in an attempt to verify whether more advanced learners' behaviour resembles that of native speakers more than less proficient learners'. One example of the structures broached in the literature would be the acquisition of pronouns and ellipsis. Research has shown that proficiency affects the acquisition of various aspects of L2 pronouns, namely, cliticisation (Scuitti, 2020), gender agreement (Dong et al., 2015) and anaphora resolution (Contemori et al., 2019). In particular, with regard to reflexive pronouns, the effect of this variable has also been confirmed in that the higher the L2 proficiency, the better the acquisition of reflexives (Yip & Tang, 1998). That is to say, higher proficient learners are able to treat the binding properties of the L2 reflexives as an independent system from their L1.

As for the literature on the interplay between proficiency, ellipsis and the interpretation of reflexives, studies have shown that higher-proficiency learners' interpretations are more similar to native speakers' than those of lower-proficiency learners (Epoge, 2012; Park, 2016; Ying, 2005). As will be shown below, Ying (2005) and Park (2016) investigated L2 English learners' interpretations of reflexives in bare, referential, and non-referential contexts, while Epoge (2012) focused only on the contrast between bare and referential contexts.

In Ying's (2005) study the participants were 50 L1 Chinese university learners of English studying English majors in Shanghai (control group: 20 native speakers of American English). They were divided into an intermediate (n=28) and an advanced (n=22) group according to a simplified version of the TOEFL testing English Structure and Reading Comprehension. The results show that in bare contexts both the

intermediate and the advanced learners interpreted the VPE sloppily, although the former did so to a greater extent (Intermediate: 82.1% sloppy/17.9% strict vs. Advanced: 74.1% sloppy/25.9% strict). These results were in line with those of the control group: 75.9% sloppy/24.1% strict reading. The differences across the two learners' groups were statistically significant, as were those between the native speakers and the intermediate learners. However, the differences between the advanced learners' group and the control one were not statistically significant. As far as referential contexts are concerned, both intermediate and advanced learners showed a preference for the strict interpretation (Intermediate: 64.9% strict/35.1% sloppy vs. Advanced: 71.1% strict/28.9% sloppy), which resembles the native speakers' behaviour (82.1% strict/17.9% sloppy). On this occasion, the differences attested among the three participant groups were statistically significant. Finally, in non-referential contexts, no such marked preferences were found between both readings in either the learners' (Intermediate: 52.5% sloppy / 47.5% strict vs. Advanced: 55.8% sloppy / 44.2% strict) or the native speakers' groups (55.6% sloppy / 44.4% strict), as the lack of statistical differences indicated.

Park (2016) replicated Ying's (2005) study with fourth-year university students in Busan (South Korea) whose proficiency level was measured using the Test of English for International Communication (TOEIC). Regarding bare contexts, his intermediate (n=13) and advanced learners (n=13) of English showed a similar preference for the sloppy interpretation (Intermediate: 91.7% sloppy/ 8.3% strict vs. Advanced: 88.5% sloppy/ 11.5% strict), which was significantly different from that of natives' (n=10; 79.2% sloppy/ 20.8% strict). As for referential contexts, native speakers slightly favoured the strict reading, intermediate learners favoured the sloppy one more clearly, and advanced learners showed nearly equal rates for either interpretation (Intermediate: 64.1% sloppy/ 35.9% strict; Advanced: 49.4% sloppy/ 50.6% strict; Native: 42.5% sloppy/57.5% strict). In this case, it is intermediate learners that statistically differed from the two other groups, no differences being found between advanced learners and native speakers. Finally, in non-referential contexts the sloppy interpretation was favoured by all of the groups, particularly native speakers and advanced learners (Intermediate: 78.2% sloppy/ 21.8% strict; Advanced: 89.1% sloppy/ 10.9% strict; Native: 94.2% sloppy/5.8% strict). Once again, the intermediate learners significantly

distanced themselves from the two other groups, no differences being observed between the control group and advanced learners.

Lastly, Epoge (2012) carried out a similar study with L2 learners of English enrolled in the Department of English at the University of Yaounde I (Cameroon) at Level One (n=40), Level Two (n= 45) and Level Three (n=43). His investigation offers the results of these three unspecified proficiency level groups (no control group of native speakers was included) with respect to only two of the contexts in point: bare and referential contexts. Regarding bare contexts, the higher the proficiency level, the more favoured the sloppy reading was (Level One: 53.1% sloppy/ 46.9% strict; Level Two: 62.5% sloppy/ 37.5% strict; Level Three: 68.6% sloppy/31.4% strict). Concerning referential contexts, the opposite tendency was observed, strict interpretations increasing as the level of proficiency raised (Level One: 21.7% sloppy/ 78.3% strict; Level Two: 13.7% sloppy/ 86.3% strict; Level Three: 8.1% sloppy/91.9% strict). Unfortunately, no inferential statistics was performed on the data.

3. Research questions

Two research questions have been entertained in an attempt to decipher the role that target language proficiency may play with regard to respondents' interpretations of English reflexive anaphora in cases of VPE. The first research question (RQ1) addresses the issue of whether target language proficiency influences participants' readings of English reflexive anaphora in VPE:

RQ1. Does target language proficiency affect the interpretation of English reflexive anaphora in VPE?

Research question 2 (RQ2) explores whether additional contextual information may have an impact on the relationship between learners' proficiency level and their interpretation of English reflexive anaphora in VPE:

RQ2. Does additional contextual information affect the relationship between learners' proficiency level and their interpretation of English reflexive anaphora in VPE?

4. Method

4.1 Participants

One hundred and four Spanish-speaking University of Cantabria students learning English and 32 native speakers of American English from North Carolina (USA) participated in this study (26 female and 6 males, average age 20.76, and age range 19-33). The non-native participants were students enrolled in BA degrees (*Early Childhood and Primary Education Teacher Training*), as well as MA students (*Second Language Teaching and Learning*). In line with Xiao's (2015) argument (see section 2.4), we used a placement test to check the proficiency levels of our students, i.e., the Quick Placement Test (Oxford University Press). The results indicated that the participants' level of English ranged between A2 and B2, according to the Common European Framework of Reference (CEFR). According to the Oxford Placement Test (OPT) assessing guidelines, a score of 16-23 meant an A2 proficiency level, one of 24-30 a B1 level, and one of 31-40 a B2 level. Three different proficiency groups could be distinguished among the participants of this study. The first group was composed by 36 A2-level students (28 females and 8 males; average age: 22.31; age range: 18-35), whose OPT mean score was 20.44 (13-23). The second group was formed by 37 B1-level students (26 females and 11 males; average age: 21.73; age range: 19-37) and whose OPT mean score was 26.65 (24-30). The third group was composed by 31 B2-level students (25 females and 6 males; average age: 23.19; age range: 19-51) and and OPT mean score of 32.90 (30-39)). These data are summarised in Table 1.

Table 1. Participants' information regarding proficiency level, gender, mean age, and OPT mean score

Proficiency level	Gender	Mean Age	OPT mean score
A2	28 Female 8 Male	22.31 (18-35)	20.44 (13-23)
B1	26 Female 11 Male	21.73 (19-37)	26.65 (24-30)
B2	25 Female 6 Male	23.19 (19-51)	32.90 (30-39)
NSs	26 Female 6 Male	20.76 (19-33)	37.67 (36-40)

4.2 Instruments

In this study we have used two types of instruments: an English proficiency test measuring the independent variable of the study, namely, target-language proficiency; and two judgement tasks to probe the readings of reflexive pronouns in cases of VPE.

In the field of applied linguistics, the proficiency determinants used can be classified into four main types: standardised tests scores, course, or grade levels, length of formal instruction, and duration of residence in the target language community. Of all these, according to Xiao (2015), the standardised test scores are among the most reliable indicators of proficiency as they allow for comparison across test takers. In this study we chose the Quick Placement Text (OUP) in its forty-item version, which measures students' proficiency up to B2 level (CEFR).

In order to test the subjects' interpretation of reflexive pronouns in cases of VPE, they had to reply to two judgement tasks based on previous works by Ying (2005), Epoge (2012), and Park (2016). The first judgement task was designed to check whether respondents show a preference for strict or sloppy interpretations of reflexive anaphora in cases of VPE with bare contexts. This task included 10 experimental sentences and 20 distractors. Example (19) illustrates an experimental sentence where the participants had to choose one of the options given immediately below.

(19) Fred talked about himself, and Robert did too.

_____ Robert talked about Robert.

_____ Robert talked about Fred.

The second task addressed the second research question mentioned above, i.e., whether the presence of a referential or non-referential context (see examples (20) and (21) respectively) affects the relationship between learners' proficiency level and their interpretation of English reflexive anaphora in VPE. This second task contained 20 experimental sentences and 20 distractors. Crucially, note that these 20 experimental sentences include the very same experimental sentences as in the first judgement task, but this time with additional referential and non-referential contexts.

(20) Fred talked about himself, and Robert did too. Robert thinks that Fred is a very interesting person.

_____ Robert talked about Robert.

_____ Robert talked about Fred.

(21) Fred talked about himself, and Robert did too. Robert went to the dentist afterwards.

_____ Robert talked about Robert.

_____ Robert talked about Fred.

4.3 Data Gathering Procedure

The data were gathered in two sessions during the students' regular class time. In the first session, participants were asked to fill in a short biographical and linguistic survey which contained questions regarding their age, gender, years of English learning, and other languages known. Subsequently, students had to respond to the Quick Placement Test (OUP), which provided information about their English language proficiency level. They completed the 40 multiple-choice questions in 30 minutes approximately. Immediately afterwards, participants were presented with the first judgement task, which they completed in around 10 minutes on average. In the second session, which took place a week later, students completed the second judgement task, which lasted around 20 minutes. As in previous research (Ying, 2005), in both judgement tasks participants were asked to select one of the two interpretations provided which matched

their initial interpretation of the underlined part of the sentence. In other words, the students were required to choose the answer that came to their minds first, that is, following their initial intuition and without going back or making any changes after their initial choice.

5. Data analysis procedure

Data were analysed quantitatively by means of both descriptive and inferential statistical procedures. As for descriptive statistics, we calculated the mean scores and standard deviations for the interpretation of English reflexive pronouns in VPE in each of the four participant groups, that is, the three proficiency learner groups (A2, B1, and B2) and the native speaker group (NS). These figures were calculated for each of the three contexts separately – 10 experimental sentences in bare contexts, 10 experimental sentences in referential contexts, and 10 experimental sentences in non-referential contexts.

Various inferential statistical analyses were performed, for all of which alpha levels of .05(*), .01(**) and .001(***) were considered as significant. Kolmogorov-Smirnov tests indicated that the distributions of the samples were skewed on nearly all occasions, so non-parametric procedures were selected to compute both inter-group and intra-group comparisons. Regarding the former, Kruskal-Wallis one-way analyses of variance were computed to verify if there were significant differences among the participant groups (A2, B1, B2, and NS), as well as post-hoc Mann-Whitney tests for all pertaining binary comparisons (A2 vs. B1, A2 vs. B2, A2 vs. NS, B1 vs. B2, B1 vs. NS, and B2 vs. NS). As for intra-group comparisons, the Friedman tests of differences among repeated measures looked for significant differences among the three contexts (bare, referential, and non-referential), whereas post-hoc paired samples Wilcoxon signed-rank tests did so in the case of the corresponding binary comparisons (sloppy vs. strict, bare vs. referential, bare vs. non-referential, referential vs. non-referential). For all binary comparisons performed, apart from calculating the statistical probability, effect sizes were computed by means of Cohen's *d*. Please note that the strength of the differences found is considered medium if above .5 and large if above .8.

6. Results

The results will be presented according to the two research questions of the study. First, the intergroup comparisons will be shown in order to discover whether target language proficiency affects the interpretation of English reflexive anaphora in VPE. Second, the intragroup comparisons will be displayed so as to explore the role of additional contextual information in the relationship between target language proficiency and the interpretation of English reflexive pronouns in VPE.

6.1 Intergroup comparisons

Regarding intergroup comparisons, the Kruskal-Wallis one-way analyses of variance revealed that there were significant differences among the four participant groups for bare ($H=10.669$; $p=.014^*$), non-referential ($H=18.457$, $p=.000^{***}$), and referential ($H=15.446$, $p=.001^{**}$) contexts. Consequently, all pertaining binary comparisons (A2 vs. B1, A2 vs. B2, A2 vs. NS, B1 vs. B2, B1 vs. NS, and B2 vs. NS) were subsequently computed through post-hoc Mann-Whitney tests for each of the contexts.

Table 2. VPE reflexive anaphora interpretation in bare contexts

	BARE CONTEXTS		
	sloppy \bar{x}	strict \bar{x}	Standard Deviation
A2 (n=36)	6.64	3.36	2.64
B1 (n=37)	7.60	2.40	2.78
B2 (n=31)	7.46	2.74	1.77
NS (n=32)	8.44	1.56	1.54

Regarding the bare contexts (Table 2), significant differences were discovered when the NS group was compared to the A2 ($z=-3.124$, $p=.002^{**}$, $d=.833$) and the B2 ($z=-2.317$, $p=.020^*$, $d=.591$) learners. As can be observed, the native speakers were the ones who interpreted English reflexives sloppily the most, whereas the A2 learners were those who did so to a lesser extent. B1 and B2 learners yielded similarly intermediate values. The B1 group was the only one who performed as the NS group, as no statistical

significance was reached in their comparison. As for the binary contrasts among learners, no statistically significant differences were found in any case.

Table 3. VPE reflexive anaphora interpretation in referential contexts

	REFERENTIAL CONTEXTS		
	sloppy \bar{x}	strict \bar{x}	Standard Deviation
A2 (n=36)	3.00	7.00	3.12
B1 (n=37)	2.73	7.27	2.83
B2 (n=31)	2.87	7.13	2.79
NS (n=32)	0.78	9.22	1.31

As far as the referential contexts are concerned (Table 3), the NS group significantly differed from the A2 ($z = -3.726$, $p = .000^{***}$, $d = .927$), the B1 ($z = -3.426$, $p = .001^{**}$, $d = .884$), and the B2 ($z = -3.656$, $p = .000^{***}$, $d = .959$) group. Native speakers clearly favoured strict readings of English reflexives when a referential context was provided. This very same tendency was found in the three learners' groups, but in this case, it was less marked than in the NS group. The lowest mean score was obtained by the A2 group, even though the B1 and B2 groups' means were just a little bit higher. In fact, no statistical differences were found in any of the binary comparisons among the learner groups.

Table 4. VPE reflexive anaphora interpretation in non-referential contexts

	NON-REFERENTIAL CONTEXTS		
	sloppy \bar{x}	strict \bar{x}	Standard Deviation
A2 (n=36)	5.80	4.20	2.55
B1 (n=37)	6.50	3.50	2.51
B2 (n=31)	6.45	3.55	3.09
NS (n=32)	8.12	1.88	2.17

With regard to the non-referential contexts (Table 4), it was the NS group the one who achieved the highest rate of sloppy interpretations. As for learners, the A2 group was the one whose mean score for sloppy readings was the lowest, the B1 and the B2 groups' means being very similar and in between the A2 and the NS groups. The inferential

analyses indicated that differences were significant when the NS group was involved in the comparison. In other words, the native speakers' behaviour was statistically different from the of A2 ($z = -3.878$, $p = .000^{***}$, $d = .980$), the B1 ($z = -2.801$, $p = .005^{**}$, $d = .690$), and the B2 ($z = -2.265$, $p = .012^*$, $d = .625$) learners. Yet, the differential behaviour of the three learner groups yielded no statistically significant differences.

6.2 Intragroup Comparisons

As for intragroup comparisons, the Friedman tests of differences indicated that there were significant differences among the three contexts (bare, referential, and non-referential) for the A2 ($\chi^2 = 42.895$, $p = .000^{***}$), B1 ($\chi^2 = 83.727$, $p = .000^{***}$), B2 ($\chi^2 = 51.776$, $p = .000^{***}$), and NS ($\chi^2 = 122.704$, $p = .000^{***}$) groups. Hence, post-hoc paired samples Wilcoxon signed-rank tests were performed for the corresponding binary comparisons to be made in each participant group.

Table 5. NSs' interpretation of VPE reflexive anaphora in the three contexts

BARE	REFERENTIAL	NON-REFERENTIAL
sloppy \bar{x} / strict \bar{x} (SD)	sloppy \bar{x} / strict \bar{x} (SD)	sloppy \bar{x} /strict \bar{x} (SD)
8.44 / 1.56 (1.54)	0.78 / 9.22 (1.31)	8.12 / 1.88 (2.17)

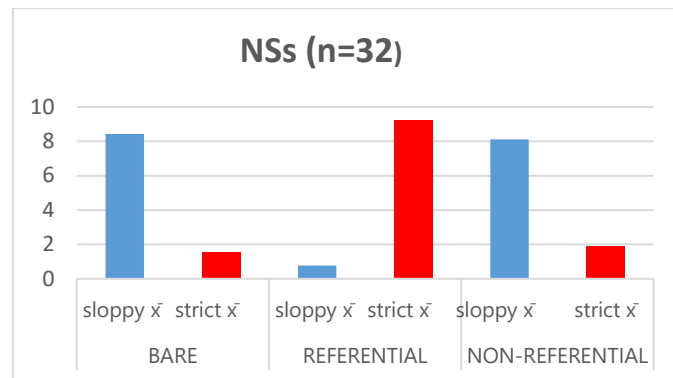


Figure 1. NSs' interpretation of VPE reflexive anaphora in the three contexts

First, we will present the results of the NS group as a baseline to which the results of the learner groups will be compared (Table 5). The native speakers' mean of the sloppy interpretation was significantly higher than the mean of the strict interpretation in both bare ($z=-4.828$, $p=.000^{***}$, $d=4.468$) and non-referential ($z=-4.536$, $p=.000^{***}$, $d=2.876$) contexts. However, the opposite tendency was found in the case of referential contexts, where the strict interpretation mean score was significantly superior to that of the sloppy interpretation ($z=-5.053$, $p=.000^{***}$, $d=6.443$). Besides, native speakers' choice of sloppy (or strict) readings in referential contexts was significantly different from native speakers' very same interpretation in bare ($z=-4.955$, $p=.000^{***}$, $d=5.358$) and non-referential ($z=-4.952$, $p=.000^{***}$, $d=4.095$) contexts, whilst no significant differences were found when the bare context interpretations were compared to the same readings in the non-referential context. In other words, NSs' preference for the sloppy interpretation prevailed in bare and non-referential contexts to the same extent, but it significantly differed from their lower preference for sloppy readings in referential contexts.

Table 6. A2 learners' interpretation of VPE reflexive anaphora in the three contexts

BARE	REFERENTIAL	NON-REFERENTIAL
sloppy \bar{x} / strict \bar{x} (SD)	sloppy \bar{x} / strict \bar{x} (SD)	sloppy \bar{x} /strict \bar{x} (SD)
6.64 / 3.36 (2.64)	3.00 / 7.00 (3.12)	5.80 / 4.20 (2.55)

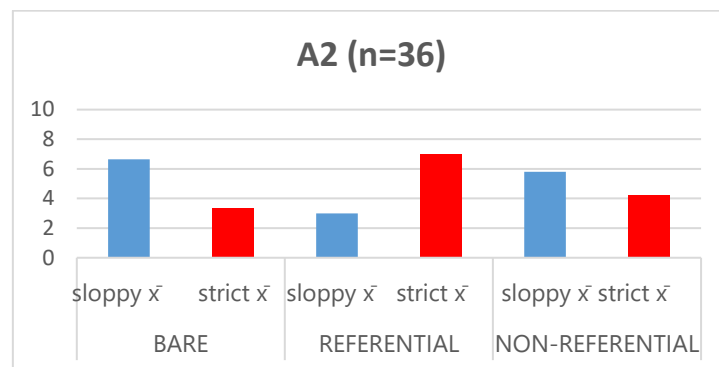


Figure 2. A2 learners' interpretation of VPE reflexive anaphora in the three contexts

Regarding the A2 learners (Table 6), the comparison between the sloppy choice and the strict choice was statistically significant in two of the contexts, namely the bare ($z=-3.305$, $p=.002^{**}$, $d=1.242$) and the referential ($z=3.149$, $p=.002^{**}$, $d=1.282$) context. As shown by the mean scores, A2 learners showed a preference for sloppy readings in bare contexts but for strict readings in referential contexts. As for non-referential contexts, the comparison between sloppy and strict choices did not reach statistical significance. When the mean score of the sloppy (or strict) interpretation in referential contexts was compared to the mean of the very same interpretation in bare ($z=-4.305$, $p=.000^{***}$, $d=1.260$) or non-referential ($z=-4.345$, $p=.000^{***}$, $d=.983$) contexts, the differences reached significance. Nevertheless, no significant differences were observed when the comparison was established between the bare and the non-referential context. That is to say, A2 learners' mean of sloppy readings in both bare and non-referential contexts was significantly higher than the mean of sloppy readings in referential contexts. Nevertheless, no differences existed between bare and non-referential contexts as far as A2 learners' interpretations since they equally favoured the sloppy reading in both.

Table 7. B1 learners' interpretation of VPE reflexive anaphora in the three contexts

BARE	REFERENTIAL	NON-REFERENTIAL
sloppy \bar{x} / strict \bar{x} (SD)	sloppy \bar{x} / strict \bar{x} (SD)	sloppy \bar{x} /strict \bar{x} (SD)
7.60 / 2.40 (2.78)	2.73 / 7.27 (2.83)	6.50 / 3.50 (2.51)

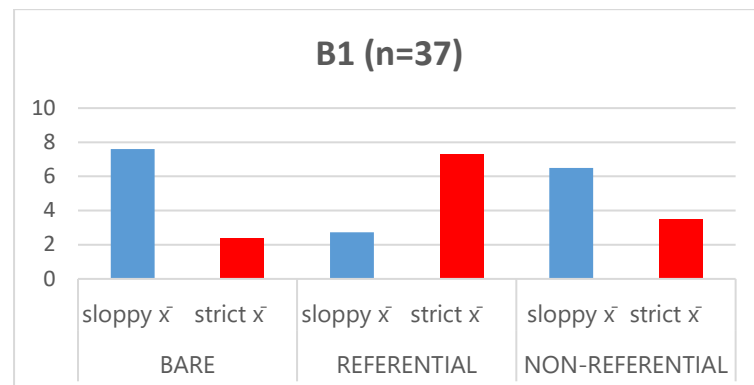


Figure 3. B1 learners' interpretation of VPE reflexive anaphora in the three contexts

As for the B1 learners (Table 7), the comparison between the sloppy and the strict readings was always significantly different, a preference for sloppy interpretations being evinced in sentences with bare ($z=-4.381$, $p=.000^{***}$, $d=1.871$) and non-referential ($z=-3.093$, $p=.000^{***}$, $d=1.195$) contexts, whereas strict interpretations prevailed in sentences with referential contexts ($z=-3.489$, $p=.000^{***}$, $d=1.604$). As for contextual differences, all types of contrasts yielded statistical significance. In other words, B1 learners' choice of sloppy (or strict) readings was significantly different when the referential context was contrasted with the bare ($z=-4.872$, $p=.000^{***}$, $d=1.856$) and the non-referential ($z=-4.772$, $p=.000^{***}$, $d=1.409$) context, but also when the bare and the non-referential contexts were compared ($z=2.192$, $p=.030^*$, $d=.415$), although in the latter the effect size was not very large. Put differently, B1 learners' mean of sloppy readings in bare contexts was statistically superior to that very same mean in either referential or non-referential contexts.

Table 8. B2 learners' interpretation of VPE reflexive anaphora in the three contexts

BARE	REFERENTIAL	NON-REFERENTIAL
sloppy \bar{x} / strict \bar{x} (SD)	sloppy \bar{x} / strict \bar{x} (SD)	sloppy \bar{x} /strict \bar{x} (SD)
7.46 / 2.74 (1.77)	2.87 / 7.13 (2.79)	6.45 / 3.55 (3.09)

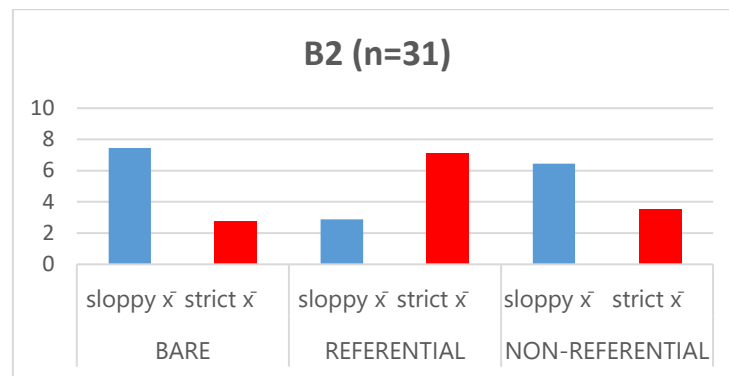


Figure 4. B2 learners' interpretation of VPE reflexive anaphora in the three contexts

Finally, the B2 learners' data (Table 8) indicated that the difference between the two types of interpretations was statistically significant in bare ($z=-4.289$, $p=.000^{***}$, $d=2.267$), referential ($z=-3.267$, $p=.001^{**}$, $d=1.527$) and non-referential ($z=-2.415$,

$p=.016^*$, $d=.939$) contexts. B2 learners also exhibited higher average scores for sloppy readings in bare and non-referential contexts, and for strict readings in referential contexts. When contextual differences were explored, inferential analyses found out statistical significance when referential vs. bare ($z=-4.682$, $p=.000^{***}$, $d=1.965$) and referential vs. non-referential ($z=-4.133$, $p=.000^{***}$, $d=1.216$) sentences were contrasted, but also when bare and non-referential contexts were compared ($z=-2.215$, $p=.027^*$, $d=.401$), even though in the latter the effect size was not large. B2 learners opted for sloppy interpretations to a larger extent in bare and non-referential contexts than in referential ones. The sloppy choice was also significantly more marked in bare than in non-referential contexts in B2 learners.

7. Discussion

We turn to the discussion of the results according to the two research questions of the study that tackle the effects of proficiency and pragmatic context.

7.1 Proficiency effect

Regarding RQ1 (Does target language proficiency affect the interpretation of English reflexive anaphora in VPE?), the data involving bare contexts confirmed that all participant groups favoured the sloppy reading. This tendency is in agreement with Principle A of the Binding Theory (Chomsky, 1981) because the anaphor in the ellipsis site is locally bound on most occasions. The fact that both readings were attested in our data offers empirical support for Sag's (1976), Kitagawa's (1991), Fiengo and May's (1994), Hestvik's (1995), and McKillen's (2016) positions, who claimed that reflexives could give rise to both strict and sloppy interpretations.

As for the intergroup comparisons among the learner groups, no statistically significant differences were attested, even though the descriptive statistics showed that the group with the lowest proficiency (A2) chose the sloppy interpretation to a lesser extent, in line with Epoge's (2012) results. This tendency contradicts Ying's (2005) and Park's (2016) findings, since their intermediate learners selected the sloppy

interpretation on more occasions than the advanced ones, a choice that even reached statistical significance in Ying's (2005) study. These divergent results might be accounted for by the way in which the proficiency variable has been operationalised in these works, since each study has considered different proficiency ranges. In the present paper, for example, we lack a group of advanced students. This fact is of particular relevance, as it may explain a more homogeneous behaviour of the overall sample of learners. Moreover, the proficiency levels observed in all these works have been measured with different tests, which may have had an impact on the way in which students were grouped into the various proficiency levels in each study.

Concerning the comparison between NSs and learners, in our data the former were the ones who chose sloppy readings in bare contexts in the vast majority of cases, a preference that significantly differed from that of A2 and B2 learners. These results do not mirror those of prior research, where NSs chose the sloppy interpretation to a significantly lesser extent than intermediate (Park, 2016; Ying, 2005) and advanced learners (Park, 2016). In light of the lack of information regarding some potential intervening variables in the studies by Ying (2005) and Park (2016), we hypothesise that factors such as the nature of the different control groups regarding their knowledge of the learners' L1, length of stay in the learners' country, knowledge of other languages, etc. might explain this discrepancy in the results.

7.2 Context effect

Regarding RQ2 (Does additional contextual information affect the relationship between learners' proficiency level and their interpretation of English reflexive anaphora in VPE?), our data regarding referential contexts show that all participant groups favoured the strict interpretation. This choice is significantly different from VPE interpretations in bare and non-referential contexts where sloppy interpretations were chosen, as indicated by the intragroup comparisons.

This finding is in line with Principle B of the Binding Theory (Chomsky, 1981), which claims that a pronominal must be non-locally bound, thus allowing for the existence of strict readings. This supports Kitawaga's (1991) and Fiengo and May's (1994) syntactic theories, whereby the reflexive acting as the antecedent can be

reconstructed as a pronoun in the ellipsis site in certain contexts. However, in the light of our data, Hestvik's (1995) claim that strict readings in VPE are disfavoured by the coordination of the antecedent and the elided clauses is not confirmed, given that all of the experimental sentences in the present study only contained coordinated clauses. Our findings would, therefore, confirm McKillen's (2016) position on the reformulation of the Condition A of the Binding Theory whereby strict readings and not only sloppy ones are available in instances of coordination if the right context is provided. Crucially, pragmatic information is offering such a context in our data, overruling the effect of syntax.

As for the various learner groups' performance in referential contexts, they all similarly selected the strict reading on nearly three quarters of the occasions, with no statistically significant differences among them. This indicates that there is no proficiency effect on the choice of interpretations in referential contexts, contrary to the results obtained in previous literature, which attested a significantly greater proportion of strict readings as proficiency increased (Epoge, 2012; Park, 2016; Ying, 2005). This divergence, once again, could be explained by the fact that our sample did not contain a group of advanced learners. However, our data coincide with those reported in the literature in that the control groups' preference for the strict reading was significantly higher than that of learners (Park, 2016; Ying, 2005). This seems to indicate that learners are less congruent than NSs in their interpretations of reflexives in VPE when they face the syntax-pragmatics interface, which poses an additional cognitive load (Gandón Chapela & Gallardo del Puerto, 2019; Sorace & Filiaci, 2006).

Lastly, in non-referential contexts our results indicated that all participant groups favoured the sloppy readings, which is in line with prior research (Park, 2016; Ying, 2005) and with the results found for the bare contexts in this study. This implies that, once again, Principle A of the Binding Theory (Chomsky, 1981) is operative (McKillen 2016) and the reflexive pronoun in the ellipsis site is locally bound on most occasions.

As for learner group comparisons, our data confirm Ying's (2005) lack of significant differences among learners, contradicting Park's (2016) results, who found that the higher the proficiency, the greater the rate of sloppy interpretations in non-referential contexts. However, it should be noted that intragroup comparisons informed that the A2 group's preference for the the sloppy readings was less marked than in

higher proficiency groups in our study. This seems to indicate that lower proficiency learners' interpretations might be misled by the pragmatic information provided to a larger extent.

Regarding the comparison between NSs and learners in non-referential contexts, the former's proportion of sloppy readings was significantly greater than that of any of the learner groups investigated. This result is in line with that reported in Park (2016), whose control group also significantly favoured the sloppy reading to a greater extent when compared with the intermediate learners (but not with the advanced ones, a group which is missing in our sample). As happened in referential contexts, this finding may be indicative of a stronger effect of the cognitive load for learners (Gandón Chapela & Gallardo del Puerto, 2019). The additional pragmatic information provided in non-referential contexts led them to a less homogeneous interpretation as a group. However, our results and Park's (2016) differ from those reported in Ying (2005), where both NSs and learner groups exhibited a slight preference for the sloppy reading. Once again, non-controlled variables such as the characteristics of the control group or of the learners' L1 could be responsible for this difference in the results.

The intragroup comparisons in non-referential contexts, which comprise data that had not been tackled statistically in previous literature, shed light on the fact that in non-referential contexts B1 and B2 learners' choice of the sloppy reading significantly decreased from their very same choice in bare contexts. In the case of A2 learners, their preference for the sloppy reading in both bare and non-referential contexts was the lowest, distancing themselves from the NSs the most. These findings contrast with NSs' similar choice in both types of contexts. Once again, this implies that the presence of the pragmatic information contained in non-referential contexts seems to have misled learners to a greater extent.

In spite of some of the differences attested in judgement patterns across the contexts under study, the learner data overall evince target-like preferences. This indicates that both Spanish learners of English and NSs are similarly sensitive to the effect of pragmatic contexts. This similarity in their behaviour entails that there might be mutual intelligibility between NSs of English and Spanish non-native speakers of English, which, to a certain extent, could reveal that this issue does not need excessive attention on the part of English teachers.

8. Conclusions

Our study attempted to explore the effect of L2 English proficiency on Spanish speakers' interpretations of reflexive pronouns in VPE as either strict or sloppy in bare, referential, and non-referential contexts.

As has been pointed out in the discussion section, the results found in previous investigations have been partially confirmed. Our study did not find substantial differences among learners even though the behaviour of the lowest proficiency group (A2) was the most distant from NSs'. Overall, in the light of the comparison established between the results of our study and prior research, what can be gathered is that there exists a high level of variability between the choices made by the participant groups in the various contexts under analysis. We hypothesise that factors such as the learners' L1, the range of the proficiency levels observed (measured with different tests), and the nature of the control group (regarding knowledge of the learners' L1, length of stay in the learners' country, knowledge of other languages) could account for this divergence in the findings. This implies that the results on the effects of proficiency over reflexive anaphora resolution in VPE contexts are not conclusive yet. Therefore, this calls for further scholarly inquiry.

In future stages, a more in-depth analysis of individual behaviour in our data would be desirable, since it might be the case that for some individual learners both sloppy and strict readings may not be available. Thus, conducting an additional independent task, such as a truth-value judgement task (Crain & Thornton, 1998) or think-aloud protocol (Yoshida, 2008), could help to shed light on this issue.

It would also be advisable to increase our sample by including Spanish advanced students of English so that the proficiency variable can be captured in its full range. In addition, it would be convenient to further explore whether speakers with different L1s vary in their interpretations of English reflexive anaphora in the various contexts under analysis. This type of research would help to disentangle whether it is UG or transfer that leads the speakers' interpretations (e.g., Al Kafri, 2013).

It is also important to note that the present paper has replicated those existing L2 studies on the interpretation of reflexives in VPE. In future work, it would be crucial to examine the variables that have already been broached in L1 studies, such as the effect

that discourse connectives (Frazier & Clifton, 2006; Hestvik, 1995; Kehler, 2000; Kim & Runner, 2009; McKillen, 2016; Ong & Brasoveanu, 2014), verb semantics (Dalrymple et al., 1991; Ong & Brasoveanu, 2014), syntactic configuration (Frazier & Clifton, 2006; Hestvik 1995; Kim & Runner, 2009; McKillen, 2016), or negation (Ong & Brasoveanu, 2014) may exert on the readings of reflexives in cases of VPE. The study of all the aforementioned variables opens up new research avenues in the field of SLA.

References

- Aelbrecht, L. (2009). *You Have the Right to Remain Silent. The Syntactic Licensing of Ellipsis* [Unpublished doctoral dissertation]. Catholic University of Brussels.
<https://lirias.kuleuven.be/retrieve/394416>.
- Aelbrecht, L. (2010). *The syntactic licensing of ellipsis*. John Benjamins.
- Al-Kafri, A. (2013). *Interpretation of English Reflexives by Child and Adult L2 Learners* [Doctoral dissertation, Newcastle University]. CORE.
<https://core.ac.uk/download/pdf/153777715.pdf>
- Al-Thubaiti, K.A. (2018). Selective vulnerability in very advanced L2 grammars: evidence from VPE constraints. *Second Language Research*, 1–28.
<https://doi.org/10.1177/0267658317751577>
- Bennett, S. (1994). Interpretation of English reflexives by adolescent speakers of Serbo-Croatian. *Second Language Research*, 10(2), 125–156.
- Bouchard, D. (1984). *On the content of empty categories*. Foris.
- Chomsky, N. (1981). *Lectures on Government and Binding*. Foris.
- Contemori, C., Asiri, O., & Perea Irigoyen, E.D. (2019). Anaphora resolution in L2 English: An analysis of discourse complexity and cross-linguistic interference. *Studies in Second Language Acquisition*, 1–28. <https://doi.org/10.1017/S0272263119000111>
- Crain, S., & Thornton, R. (1998). The truth value judgment task: Fundamentals of design. *University of Maryland working papers in linguistics*, 6, 61–70.
- Dalrymple, M. (2005). Against reconstruction in ellipsis. In R. Elugardo & R.J. Stainton (Eds.), *Ellipsis and nonsentential speech. Studies in Linguistics and Philosophy*, vol. 81 (pp. 31–55). Springer. https://doi.org/10.1007/1-4020-2301-4_2

- Dalrymple, M., Shieber, S. M., & Pereira, F. C. (1991). Ellipsis and higher-order unification. *Linguistics and philosophy*, 14(4), 399–452.
- Dong, Y., Wen, Y., Zeng, X., & Ji, Y. (2015). Exploring the cause of English pronoun gender errors by Chinese learners of English: Evidence from the self-paced reading paradigm. *Journal of Psycholinguistic Research*, 44(6), 733–747.
- Duffield, N.G., & Matsuo, A. (2009). Native speakers' versus L2 learners' sensitivity to parallelism. *Studies in Second Language Acquisition*, 31, 93–123.
- Epoge, N. K. (2012). Reflexive anaphora in VP-elliptical sentences of ESL learners in Cameroon. *Syllabus Review*, 3(1), 257–279.
- Fiengo, R., & May, R. (1994). *Indices and identity*. MIT Press.
- Frazier, L., & Clifton C. (2006). Ellipsis and discourse coherence. *Linguistics and Philosophy*, 29(3), 315–46.
- Gandón Chapela, E., & Gallardo del Puerto, F. (2019). Verb phrase ellipsis and reflexive anaphora resolution in second language acquisition: A study of Spanish learners of English. *Miscelánea: A Journal of English and American Studies*, 59, 71–94.
- Hankamer, J., & Sag, I. A. (1976). Deep and surface anaphora. *Linguistic Inquiry*, 7, 391–428.
- Hestvik, A. (1995). Reflexives and ellipsis. *Natural Language Semantics*, 3, 211–237.
- Hirakawa, M. (1990). A study of the L2 acquisition of English reflexives. *Second Language Research*, 6(1), 60–85.
- Hwang, H. (2020). *A contrast between VP-Ellipsis and Gapping in English: L1 acquisition, L2 acquisition, and L2 processing*. [Doctoral dissertation, University of Hawai'i at Mānoa]. Scholar Space.
<https://scholarspace.manoa.hawaii.edu/handle/10125/69071>
- Johnson, K. (2001). What VP ellipsis can do, and what it can't, but not why. In M. Baltin & C. Collins (Eds.), *The handbook of contemporary syntactic theory* (pp. 439–79). Blackwell Publishers.
- Kehler, A. (2000). Coherence and the resolution of ellipsis. *Linguistics and Philosophy*, 23, 533–575.

- Kim, J. (2015). Korean L2 Learners' comprehension of VP-ellipsis and one-substitution in English. *Enekwahakyenkwu [The Journal of Linguistics Science]*, 75, 107–128.
- Kim, C.S., & Runner, J. (2009). Strict identity, coherence, and parallelism in VP ellipsis. In E. Cormany, S. Ito, & D. Lutz (Eds.), *Semantics and Linguistic Theory. Proceedings of the 19th Semantics and Linguistic Theory Conference* (pp. 275–287). <https://doi.org/10.3765/salt.v19i0.2535>
- Kitagawa, Y. (1991). Copying identity. *Natural Language and Linguistic Theory*, 9, 497–536.
- Lee, D., & Schachter, J. (1997). Sensitive period effects in Binding Theory. *Language Acquisition*, 6(4), 333–362.
- Lozano, C. (2021). Anaphora Resolution in Second Language Acquisition. In M. Aronoff (Ed.), *Oxford Bibliographies in Linguistics*. Oxford University Press. <https://dx.doi.org/10.1093/OBO/9780199772810-0268>.
- MacLaughlin, D. (1998). The acquisition of the morphosyntax of English reflexives by non-native speakers. In M. L. Beck (Ed.), *Morphology and its interfaces in second language knowledge* (pp.195–226). John Benjamins.
- Mart, C. T. (2018). From communicative competence to language development. *International Journal of English Linguistics*, 8(2), 163–167.
- Matsuo, A., & Duffield, N. (2001). VP-ellipsis and anaphora in child language acquisition. *Language Acquisition*, 9, 301–327.
- McKillen, A. (2016). *On the interpretation of reflexive pronouns*. [Doctoral dissertation, McGill University]. Library and Archives Canada. <https://www.bac-lac.gc.ca/eng/services/theses/Pages/item.aspx?idNumber=953107291>
- Ong, M., & Brasoveanu, A. (2014). Strict and sloppy reflexives in VP-ellipsis. In C. Piñón (Ed.), *Empirical Issues in Syntax and Semantics 10* (pp. 251–268). <http://www.cssp.cnrs.fr/eiss10/>
- Park, E. (2016). Korean English learners' interpretation of reflexive anaphora in VP-ellipsis. *Journal of Language Sciences*, 23(3), 111–135.
- Postman, W.A., Foley, C., Pactovis, J., Rothenstein, B., Kaye, M., Lowe, D., & Lust, B. (1997). Children's knowledge of verbal inflection and LF raising: New evidence from elicited production of VP ellipsis structures. In S. Somashekar, K.

- Yamakoshi, M. Blume, & C. Foley (Eds.), *Cornell working papers in linguistics: Papers on language acquisition 15* (pp. 39–64). Cornell Linguistics Circle.
- Roeper, T. (2018). Ellipsis and acquisition. In J. van Craenenbroeck & T. Temmerman (Eds.), *The Oxford handbook of ellipsis* (pp. 276–313). Oxford University Press.
- Ross, J. R. (1967). Constraints on variables in syntax. [Doctoral dissertation, MIT]. MIT Libraries. <https://dspace.mit.edu/handle/1721.1/15166>
- Sag, I. A. (1976). Deletion and Logical Form [Doctoral dissertation, MIT]. MIT Libraries. <https://dspace.mit.edu/handle/1721.1/16401>
- Scuitti, S. (2020). The acquisition of clitic pronouns in complex infinitival clauses by German-speaking learners of Italian as an L3: The role of proficiency in target and background language(s). In C. Bardel & L. Sánchez (Eds.), *Third language acquisition: Age, proficiency and multilingualism* (pp. 111–144). Language Science Press.
- Thornton, R., & Wexler, K. (1999). *Principle B, VP ellipsis, and interpretation in child grammar*. MIT Press.
- Williams, E. (1977). Discourse and Logical Form. *Linguistic Inquiry*, 8(1), 101–39.
- Xiao, F. (2015). Proficiency effect on L2 pragmatic competence. *Studies in Second Language Learning and Teaching*, 5(4), 557–581.
<https://doi.org/10.14746/sllt.2015.5.4.3>
- Ying, H. G. (2005). Second language learners' reading of reflexive anaphora in VP-ellipsis: A Relevance Theory perspective. *Language Sciences*, 27, 551–570.
- Yip, V. & Tang, G. (1998). Acquisition of English reflexive binding by Cantonese learners: Testing the Positive Transfer Hypothesis. In M.L. Beck (Ed.), *Morphology and its interfaces in second language knowledge* (pp. 165–193). John Benjamins.
- Yoshida, M. (2008). Think-Aloud Protocols and Type of Reading Task: The Issue of Reactivity in L2 Reading Research. In M. Bowles, R. Foote, S. Perpiñán, & R. Bhatt (Eds.), *Selected Proceedings of the 2007 Second Language Research Forum* (pp. 199–209). Cascadilla Proceedings Project.

La interpretación de los pronombres reflexivos en casos de elipsis del sintagma verbal por parte de aprendices de inglés como L2 con distintos niveles de competencia

Resumen

La anáfora reflexiva en inglés en casos de elipsis del sintagma verbal puede dar lugar a dos posibles lecturas, ya que podría interpretarse de manera estricta o laxa. Algunos estudios previos en el campo de adquisición de segundas lenguas (Epoge, 2012; Park, 2016; Ying, 2005) se han centrado en determinar el papel que la competencia lingüística en una segunda lengua puede ejercer en las interpretaciones de los aprendices de los pronombres reflexivos en oraciones sin información adicional, con contextos referenciales y con contextos no referenciales. El presente artículo proporciona datos de 104 aprendices de inglés españoles (niveles A2, B1 y B2) y 32 hablantes nativos de inglés. Los resultados muestran que los participantes tienden a interpretar los pronombres reflexivos de forma laxa en oraciones sin información adicional y con contextos no referenciales, mientras que en aquellas con contextos referenciales prevalecen las lecturas estrictas. Existen diferencias significativas en la interpretación de los aprendices con respecto a los hablantes nativos, aunque las diferencias entre los tres grupos de aprendices no son tan marcadas. Sin embargo, el grupo con menor competencia lingüística es el que más difiere de los hablantes nativos. Los hallazgos confirman parcialmente investigaciones previas y las discrepancias halladas podrían atribuirse de forma tentativa a variables externas tales como la L1 del alumnado, el rango de los niveles de competencia lingüística o las características de los grupos de control.

Palabras clave: elipsis del sintagma verbal, pronombres reflexivos, interpretación estricta, interpretación laxa, nivel de competencia lingüística, inglés como segunda lengua

Address for correspondence

Francisco Gallardo-del-Puerto

Universidad de Cantabria

Departamento de Filología

Avenida de los Castros, s/n

39500, Santander, España

francisco.gallardo@unican.es

<https://orcid.org/0000-0001-8578-9861>

Evelyn Gandón-Chapela

Universidad de Cantabria

Departamento de Filología

Avenida de los Castros, s/n

39500, Santander, España

evelyn.gandon@unican.es

<https://orcid.org/0000-0002-3999-073X>

Publication history

Date received: 10 November 2021

Date accepted: 29 December 2022