Gandón-Chapela, Evelyn & Gallardo-del-Puerto, Francisco. 2025. The effect of gender (mis)match on the interpretation of English reflexive pronouns in cases of VP-ellipsis by native and non-native speakers of English. *Glossa: a journal of general linguistics* 10(1). pp. 1–33. DOI: https://doi.org/10.16995/glossa.11493

The effect of gender (mis)match on the interpretation of English reflexive pronouns in cases of VP-ellipsis by native and non-native speakers of English

Evelyn Gandón-Chapela, Universidad de Cantabria, evelyn.gandon@unican.es

Francisco Gallardo-del-Puerto, Universidad de Cantabria, francisco.gallardo@unican.es

This paper aims to investigate the effect of both gender (mis)match and the direction of the gender mismatch on the interpretation of English reflexive pronouns in cases of VP-ellipsis by native speakers of English and L1 Spanish learners of English with different proficiency levels (intermediate, upper intermediate and advanced). Results showed that native speakers of English tend to prefer the sloppy reading over the strict one and are sensitive to gender mismatches when interpreting reflexives in VP-ellipsis constructions. Our data support a non-structural account of ellipsis resolution. Concerning learners, no gender mismatch effects were attested, since, in general, their preferred choice of sloppy interpretations was less frequent in both matched and mismatched contexts. Nevertheless, proficiency turned out to be a significant factor in the interpretation of gender-matched sentences, as intermediate learners interpreted the experimental sentences sloppily significantly less frequently than native speakers, whilst advanced learners did not. Finally, all participants showed a more marked preference for sloppy readings in feminine-masculine orderings.

1 Introduction

As is well-known in the literature on ellipsis, the interpretations of English reflexive pronouns in cases of VP-ellipsis (VPE[1] henceforth) have given rise to fruitful research in the field of theoretical linguistics in the past fifty years (e.g., Sag 1976; Dalrymple et al. 1991; Kitagawa 1991; Hestvik 1995; Fiengo & May 1994; Dalrymple 2005; McKillen 2016) due to the fact that reflexive anaphora in VPE may receive two possible readings. These two potential interpretations, known as sloppy and strict in the literature, could be explained under Principle A (sloppy reading) and Principle B (strict reading) of the Binding Theory (Chomsky 1981) and the c-commanding relations established between the proper nouns and the reflexive pronouns. For instance, in a sentence such as "Sienna criticised herself and Melanie did too", we could interpret the ellipsis site as meaning either "Melanie criticised Melanie" (sloppy reading) or "Melanie criticised Sienna" (strict reading).

To our knowledge, there has been only one L1 study (Storoshenko 2017) that has looked into the interaction between VPE, reflexive anaphora and (mis)matches. Taking the example that has just been mentioned above, a gender-matched sentence would be "Sienna criticised herself and Melanie did too", while a gender-mismatched one would be "Sienna criticised herself and George did too". Since there is also a dearth of studies investigating this interplay in English as a second language, in the present paper we will explore the effect of both gender (mis)match and the direction of the mismatch ("Sienna criticised herself and George did too" vs. "George criticised himself and Sienna did too") on the readings of reflexive pronouns in cases of VPE made by native speakers of English as well as by three different proficiency level groups (B1, B2 and C1) of L1 Spanish learners of English. Thanks to the empirical study of English L1 and L2 speakers' comprehension and interpretation of ellipsis (mis)matches, it will be possible to contribute to the (non)structuralism and identity condition theoretical debate concerning ellipsis, as our data will help to throw some light on whether the syntactic representation of the antecedent is essential for ellipsis resolution and whether sentences containing mismatching inflectional features may have an impact on the interpretation of ellipsis.

[1] We are aware of the fact that, as Miller & Pullum (2014: 6) put forward, the name 'VP ellipsis' is a "very poorly chosen term, [. . .] because it is neither necessary nor sufficient that it should involve ellipsis of a VP". Following Sag (1976: 53), Miller & Pullum (2014: 6) claim that this type of construction should be called 'Post-Auxiliary Ellipsis', since "the defining characteristic is not that a VP is omitted but that a constituent or constituent sequence immediately following an auxiliary is missing" (Miller & Pullum 2014: 6). In the present paper we use the misnomer VPE because it is the prevalent one in the ellipsis literature on theoretical syntax (e.g., Hankamer & Sag 1976; Sag 1976; Williams 1977; Hardt 1993; Lobeck 1995; Johnson 2001; Merchant 2008; Aelbrecht 2010) and psycholinguistics (e.g., Garnham et al. 1998; Frazier & Clifton 2005; 2006; Arregui et al. 2006; Clifton & Frazier 2010; Kim et al. 2011; Runner & Head 2014), as well as the one used in previous research on the same phenomenon in the field of Second Language Acquisition (Ying 2005; Epoge 2012; Park 2016; Gandón-Chapela & Gallardo-del-Puerto 2019). Therefore, following Warner (1993: 5–6), the misnomer VPE has been treated as a subtype of Post-Auxiliary Ellipsis that can omit both verbal and non-verbal material (see also Gandón-Chapela 2016; 2020a).

Regarding the previous literature on the interpretation of English reflexive anaphora in cases of VPE without taking into account gender (mis)matches, this paper comes to fill in the gap by adding a group of L1 Spanish advanced learners of English, as lower proficiency levels of this type of learners have already been investigated (Gallardo-del-Puerto & Gandón-Chapela 2024).

As for our look into L2 data, in line with Duffield & Matsuo (2009), it should be noted from the very beginning that our main purpose is to analyse and discuss English L2 learners' data to help arbitrate among competing explanations of ellipsis resolution and the identity condition on ellipsis for any group of learners, apart from native speakers. Nevertheless, as different learner proficiencies are explored, the results reported in this paper may shed light on relevant issues in SLA such as the pace of learning and eventual attainment (Ying 2005; Epoge 2012; Park 2016).

In this paper, section 2 provides an overview of the state of the art on the interplay between reflexive anaphora, gender (mis)matches, ellipsis and language proficiency. In section 3, we will describe the three research questions under scrutiny. In section 4 we will explain the procedure followed to gather and analyse the data. Section 5 presents the results, which are subsequently discussed in section 6. Section 7 offers a summary of the main findings, concluding remarks, and issues for further research.

2 State of the art

2.1 VPE

As is well-known, VPE is a special kind of ellipsis in the sense that it is less constrained than other elliptical constructions such as gapping or sluicing. For instance, (i) it may be exophoric (as in example (1), see Miller & Pullum's 2014 seminal work), (ii) it may be anaphoric or cataphoric (as in examples (2) and (3), respectively), (iii) it may allow for voice mismatches between the antecedent and the ellipsis site (as in example (4), see Kertz 2013; Poppels & Kehler 2019; Gandón-Chapela 2020a; 2020b), and (iv) it may allow for different inflectional features with respect to its antecedent (as in example (5)):

(1) [to a child in front of a hot oven]:
– Don't!

(2) If you want me to invite Kim as well, I will ~~invite Kim as well~~.[2] (Huddleston & Pullum (2002: 1456)

(3) If you want me to ~~invite Kim as well~~, I will invite Kim as well. (Huddleston & Pullum (2002: 1456)

[2] Struck-out words represent elided material. Strikethrough is merely used as an expository device in this paper, as no syntactic or semantic analysis is intended.

(4) The incident was reported by the driver, although he didn't really need to ~~report the incident~~. (Kertz 2013: 390)

(5) Alice should have said it, but she did not ~~say it~~.

It should also be noted that certain discourse conditions have been shown to play a role in cases of VPE, whose focus may be either auxiliary choice or subject choice. These discourse conditions have been tackled by Kehler (2000; 2002), Kertz (2010; 2013), Miller (2011; 2014), Miller & Pullum (2014), or Miller & Hemforth (2014). Following Kertz (2008), Miller (2011) and Miller & Pullum (2014), we have distinguished two central uses of Post-Auxiliary Ellipsis (PAE henceforth, where VPE would be a subtype of PAE): auxiliary choice and subject choice, which would need to comply with the following conditions (Miller & Pullum 2014: 12):

> Type 1: Auxiliary choice
>
> Formal characteristics: The subject of the antecedent is identical with the subject of the PAE construction and the auxiliary is (at least weakly) stressed, signalling a new choice of tense, aspect, modality, or (in the most overwhelmingly frequent case) polarity.
>
> Discourse requirement: A choice between the members of a jointly exhaustive set of alternative situations must be highly salient in the discourse context, and the point of the utterance containing the PAE is strictly limited to selecting one member of that set.
>
> Type 2: Subject choice
>
> Formal characteristics: The subject of the antecedent is distinct from the subject of the PAE construction, and stressed if it is a pronoun.
>
> Discourse requirement: A particular property must be highly salient in the discourse context, and the point of the utterance containing the PAE must be strictly limited to identifying something or someone possessing that property.

These two different types of focus (marked in italics) are exemplified by Miller & Pullum (2014) in instances (6) and (7), respectively:

(6) A: He shops in women's.
B: No, he *doesn't*. [COCA]

(7) a. *The boys* cheered. *I* did too.
b. *She* didn't say anything, and *I* didn't either.

As can be observed, in (6) the focus is on the auxiliary (*doesn't*), marking the change of polarity in the ellipsis site with respect to its antecedent (as the subject of both clauses is the same). In the examples provided in (7), on the other hand, one can observe that the focus is placed

on the contrast of subjects between the antecedent and the target of ellipsis (i.e., *the boys* vs. *I* and *she* vs. *I*). The role played by these discourse conditions is relevant for this paper, since, as mentioned in the literature, more mismatches seem to be allowed in the auxiliary choice condition (see Kertz 2013; Miller & Pullum 2014) between the source and the target of ellipsis. This entails that gender mismatch, the object of study of this paper, could play a (small) role, given that the materials that will be tested are all subject choice, that is, the focus is placed on the contrast between the subjects of the antecedent and the ellipsis site (see section 4.2 for more details).

In the following subsection, we will provide an overview of the two questions that have been traditionally investigated in the ellipsis literature, i.e., the structure and the identity questions.

2.2 Ellipsis and gender (mis)matches

Two of the long-standing debates in the study of ellipsis revolve around trying to decipher whether there is structure at the ellipsis site that needs to be computed (known as the "structure question" in the literature) and determining the type of relation held between the elided material and its antecedent (also known as the "identity question"). In the case of the former, the two possible answers that could be given to this question have crucial consequences for the theory of grammar, since, if the answer is affirmative, the theories of grammars must posit the existence of unpronounced phrases and heads. There is a vast literature on ellipsis dedicated to arguing for some evidence that would support either the structural or the non-structural stances, and this has obviously been done indirectly: if there are effects on the surrounding material which may be attributed to elided elements — the so-called 'connectivity effects' —, then one can claim that there is syntax internal to the ellipsis site; if, on the contrary, 'non connectivity effects' between the elided material and the surrounding material are found, then there is evidence to posit that there is no syntactic structure internal to the ellipsis site, i.e., 'what you see is what you get'. In Merchant's words (2019: 25), "if effects are found which we would otherwise attribute to properties of structure X in similar, nonelliptical, cases, but structure X is, by hypothesis, internal to the ellipsis site, then X exists. If, on the other hand, expected properties are missing, one could conclude that structure X is absent". Those in favour of a structural approach (Sag 1976; Williams 1977; Fiengo & May 1994; Chung et al. 1995; Kehler 2002; Sag & Hankamer 1984; Merchant 2001; van Craenenbroeck 2010; Aelbrecht 2010; among others) have shown that there exist some connectivity effects between the antecedent and the ellipsis site (locality effects, case matching, preposition- stranding, the distribution of complementizers, infinitivals and predicate answers), while those defending a non-structural approach (known as the direct interpretation) (Keenan 1971; Dalrymple et al. 1991; Hardt 1993; Ginzburg & Sag 2000; Culicover & Jackendoff 2005; among others) have provided some counter-evidence illustrating the absence of unpronounced structure (absence of locality effects, exceptions to the P-stranding generalization, voice and

category mismatches, violation of binding principles, lack of case-matching effects, pragmatic antecedents, non-extractable remnants, etc.).[3]

Concerning the identity question, most of the debates have focused on determining the type of recoverability condition that is at stake in cases of ellipsis, since omitted elements need to be recoverable from the (either linguistic or extralinguistic) context in order for them to be elided. Hankamer & Sag (1976) proposed a distinction between deep (forms such as *do it*, *do this*, *do that*, *do so*) and surface (e.g., VPE) anaphora putting forth that the former does not require a syntactic parallelism and it can be exophoric, whereas the latter does and therefore it can never be exophoric. However, as shown above, research has revealed that exophora and voice mismatches are indeed possible with surface anaphora such as VPE (see, among others, Webber 1979; Dalrymple et al. 1991; Hardt 1993; Kehler 2000; Jacobson 2003; 2008; 2009; Kertz 2010; Miller & Pullum 2014). Traditionally, several identity relationships between the antecedent and the ellipsis site have been proposed in the literature: semantic (Dalrymple et al. 1991; Hardt 1993; Ginzburg & Sag 2000; Merchant 2001; Merchant 2004; Culicover & Jackendoff 2005; van Craenenbroeck 2010; Aelbrecht 2009; 2010; Thoms 2010; 2013), syntactic (Sag 1976; Williams 1977; Fiengo & May 1994; Chung et al. 1995) or both (Chung 2006; 2013; van Craenenbroeck 2010; Merchant 2013).[4] In this regard, research has shown that syntactic parallelism seems to play a role in constraining the relation held between the elided material and its antecedent.[5] For instance, antecedent and ellipsis morphosyntactic mismatches can result in ungrammaticality when *be* and the auxiliary *have* are involved, as observed in the examples provided by Lasnik (1999) in (8):

(8) a. *John was here, and Mary will too <be here>.[6]
b. *John has left, but Mary shouldn't <have left>.
c. John was here, and Mary too <was here>.
d. John has left, but Mary hasn't <left>.

[3] For reasons of space, we will not provide examples of the (non)connectivity effects purported in the literature, but we refer the reader to Gandón-Chapela (2016), Bîlbîie (2017), Merchant (2019) or Poppels (2022), for more details.

[4] See Poppels (2022) for a recent reanalysis of the identity and referential theories regarding ellipsis, arguing that the latter provide an answer to the theoretical questions that are typically attributed to the identity-based approach. In fact, it has been shown that VPE does not require syntactic identity in examples such as the following (Poppels & Kehler 2018, cited in Poppels 2022: 353):

A: Can I borrow your textbook over the weekend?
B: I can't ~~lend it to you~~: I'll need it myself.

[5] It should be noted that there is also evidence in favour of the fact that there may be overall processing constraints on coordination which favour parallelism independently of ellipsis (see Frazier et al. 1984), as pointed out by one of the anonymous reviewers.

[6] Please note that angle brackets illustrate elided material.

The ungrammaticality of examples (8a) and (8b) stands in contrast with the grammaticality of the matched control examples in (8c) and (8d), which demonstrates that the lack of morphosyntactic isomorphism must be the reason why (8a) and (8b) are judged unacceptable.

However, it must be noted that absence of morphosyntactic parallelism, e.g., gender mismatch, seems to be grammatical with nominal ellipsis (Merchant 2014; Sudo & Spathas 2020; Kučerová et al. 2021; Sprouse et al. 2022) and predicative ellipsis (Lasnik 1999; Depiante & Masullo 2001; Nunes & Zocca 2005; 2009; Bobaljik & Zocca 2011; Merchant 2014; Aparicio et al. 2015) in several languages such as English, Spanish, Brazilian Portuguese, Russian, German or Greek. What seems to be at stake in these languages is that inflectional features may not be relevant for ellipsis computation. For instance, Merchant (2014: 10) claims that the example from Greek illustrated in (9) is grammatical despite containing a gender mismatch between the ellipsis site and its antecedent.

(9) *I Maria ine ikani, ala o Alexandros dhen ine <ikanos>.*
the Maria is capable-FEM-SG but the Alexandros not is <capable-MASC-SG>
'Maria is capable, but Alexandros is not.'

In order to shed some light on this particular issue, Aparicio et al. (2015) tested the effect of gender (mis)matches in Spanish stripping (see example (10)), another similar construction which contains the neuter particle 'lo' (i.e., a deep anaphor, see example (11), emphasis in the original), and non-elliptical sentences thanks to two eye-tracking experiments that involved 28 and 21 L1 Spanish participants, respectively. Subjects had to perform a yes-no grammaticality judgement about the sentence they had just read.

(10) *Juan es alto y María también.*
Juan is tall-MASC-SG and María too
'Juan is tall and María too.'

(11) *Juan es alto y María **lo** es también.*
Juan is tall-MASC-SG and María **it-NEUT** is too
'Juan is tall and María is too.'

Their results showed that, in general, there was a high acceptability of gender mismatches ($\geq$80%). However, participants were highly sensitive to morphological feature mismatch, as there were degraded acceptability judgments and elevated reading times for elliptical mismatched conditions. Interestingly, these authors also found that the ordering of the gender mattered too: the ordering feminine-masculine was judged less acceptable than the ordering masculine-feminine in contexts where there was a mismatch. Aparicio et al.'s (2015) results also pointed to the fact that mismatches were judged more acceptable in cases of deep anaphora as opposed to surface anaphora (stripping) (which is in line with Hankamer & Sag 1976) and that the gender mismatch

penalty was completely absent for non-elliptical counterparts (Aparicio et al. 2015: 2). Thus, their conclusion was the following: “Elliptical sentences containing conflicting inflectional phi-features are more costly to process than elliptical sentences without feature mismatch” (Aparicio et al. 2015: 2). That is to say, morphological features would be relevant for ellipsis computation.

Abeillé & Kim (2022), on their part, conducted a corpus-based study on English and French additive fragments, a type of stripping except for the presence of an additive adverb (*too/aussi*), by analysing data after 1980 from well-edited corpora such as the COCA (Davies 2008) and Frantext (www.frantext.fr). As these authors mentioned, since Hankamer & Sag (1976), stripping had been considered a surface anaphor which was more constrained than deep anaphors. However, they challenged this idea by providing empirical data showing that additive fragments have “much more flexibilities in their uses than previously thought and even allow for non-local antecedents and exophoric uses” (Abeillé & Kim 2022: 498). The following would be an example of exophoric stripping in English, where the context functions as the non-verbal antecedent (Abeillé & Kim 2022: 511):

(12) At the Riverside Caf, Kyle orders something that sounds like chicken-fry-stick and I nod and say, “Me too.” (COCA 2019 FIC)

Crucially for the purposes of the present paper, these authors also pointed to the fact that gender mismatch is possible in French stripping (Abeillé & Kim 2022: 502), as exemplified in (13):

(13) Cette fois, je suis vraiment amoureuse, lui aussi.
this time, I.$_{CL}$ am really lover.$_{FSG}$ him too
‘This time, I am really in love, him too.’ (Frantext: Prin 2005)

Their analysis of stripping constructions in both languages led the authors to propose a direct interpretation approach (that is, a non-structural one in which ‘what you see is what you get’) along the lines of Ginzburg and Sag (2000), thus defying previous accounts based on syntactic reconstruction and coordination.

Finally, Sprouse et al. (2022) analysed gender mismatches in nominal ellipsis in order to probe the ellipsis test put forward by Bobaljik & Zocca (2011) for the gender asymmetries that seem to arise in English noun pairs such as *actor/actress*, *waiter/waitress*, etc. (asymmetric nouns) vs. *prince/princess, king/queen*, etc. (symmetric nouns), as illustrated in examples (14) and (15) (Sprouse et al. 2022: 349–350):

(14) a. John is an actor, and Mary is too.
b. *Mary is an actress, and John is too.

(15) a. *John is a prince, and Mary is too.
b. *Mary is a princess, and John is too.

As can be observed, in (14) these nouns evince an asymmetry under ellipsis due to the fact that the unmarked form (*actor*) can be the antecedent of a female elided predicate, whereas the marked form (*actress*) cannot be the antecedent of a male elided predicate. In (15), however, both noun orderings result in unacceptability. As Sprouse et al. (2022: 349) put it, "[t]he underlying idea of the ellipsis test is that the identity requirement on ellipsis can be leveraged to uncover these asymmetries, and crucially, convert those asymmetries into unacceptability". Thus, these authors administered an acceptability judgement test to 2,612 English native speakers based on 7-point scale task containing sentences such as the ones shown in (14) and (15), as well as their non-elliptical (e.g., *John is an actor and Mary is an actor too*) and gender-matched counterparts (e.g., *John is an actor and Bill is an actor too*). The results of their study showed that the expected differences between asymmetric and symmetric nouns were detected and could be accounted for without resorting to ellipsis, that is to say, ellipsis was not a requirement for mismatches to be tolerated (contra Bobaljik & Zocca 2011). It is important to note that, in the case of asymmetric noun pairs, these authors found that, independently of ellipsis, the masculine-feminine ordering (*John is an actor and Mary is too* and *John is an actor and Mary is an actor too*) was preferred over the feminine-masculine ordering (*Mary is an actress and John is too* and *Mary is an actress and John is an actress too*). This preference for masculine-feminine gender orderings would be in line with Aparicio et al.'s (2015) findings for Spanish stripping mentioned above. In addition, Sprouse et al. (2022) discovered that there was a significant increase in the acceptability of feminine-masculine orderings in elliptical constructions for two of the noun pairs under study (*queen/king* and *widow/widower*).

2.3 Reflexive anaphora and gender (mis)matches

In the past three decades, psycholinguistic research has drawn its attention towards the interplay between reflexive anaphora and VPE and the two possible interpretations that may arise: strict or sloppy (Shapiro & Hestvik 1995; Frazier & Clifton 2000; Shapiro et al. 2003; Frazier & Clifton 2006). For instance, Shapiro & Hestvik (1995) run two experiments that studied the on-line processing of coordinated and subordinated VPE containing reflexive anaphora, as illustrated in (16) and (17), respectively (Shapiro & Hestvik 1995: 517):

(16) The policeman defended himself and the fireman did [e][7] too, according to someone who was there.

(17) The policeman defended himself because the fireman did [e], according to someone who was there.

Their aim was to try to decipher, thanks to a priming technique, if listeners (n = 60 college students) automatically access the strict reading immediately after reading the bare auxiliary by

[7] [e] stands for 'ellipsis'.

showing reaccess of the subject noun phrase in the antecedent clause. Their findings showed that depending on the type of relation established between the two clauses, different time courses of activation emerged: in coordinated VPE, gap filling of the strict reading was immediate, whereas in subordination, this only occurred "later during the temporal unfolding of the sentence" (Shapiro & Hestvik 1995: 530). As Frazier & Clifton (2006) state, these findings suggest that "the strict interpretation ("the fireman defended the policeman") is computed, at least temporarily, in sentences like (16), where the strict interpretation is clearly unpreferred". Interestingly, this has been claimed to occur even with sentences such as (18), where the strict interpretation would not be possible due to the nature of the verb (Shapiro et al. 2003: 6):

(18) The policeman perjured himself and the fireman did too.

Frazier & Clifton (2006), on their part, tested Kehler's (2000; 2002) claim that a reflexive may receive a strict interpretation only if a causal discourse coherence relation is established (as in example (19a), but not with a resemblance coherence relation (as in (19b)) (Frazier & Clifton 2006: 8). In other words, according to Kehler, Principle A would need to be violated in sentences involving causal relations, but obeyed in those involving resemblance ones.

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

The prior findings reported by Shapiro & Hestvik (1995) and Shapiro et al. (2003) would suggest that the strict interpretation is computed even in sentences involving a resemblance coherence relation and even for sentences which may not receive a strict reading in the end. Frazier & Clifton (2006) found that participants (n = 48 University of Massachusetts students) responding to a forced-choice task with the two possible interpretations (strict or sloppy) chose strict readings in both causal (53%) and resemblance (48%) sentences, and the differences attested were not statistically significant, which means that "there was no substantial difference in the probability of assigning a strict interpretation to a reflexive in VP ellipsis in cause-effect relations vs. in resemblance relations" (Frazier & Clifton 2006: 8). Therefore, these findings disconfirm Kehler's (2000; 2002) hypothesis that a strict interpretation would only be licensed in causal relation sentences.

There have also been some studies that have analysed the interplay between English reflexive anaphora and gender (mis)matches. For instance, Runner & Head (2014) focused their study on how the grammatical constraints of the syntactic Binding Theory, that is, the structural constraints on reflexives and pronouns, apply during on-line processing. They used a novel visual world eye-tracking method which manipulates the gender of potential antecedents visually and they found clear evidence that native speakers of English (n = 25) consider gender-matching potential antecedent NPs for reflexives and pronouns that match in gender regardless of whether

they are licensed structurally by the Binding Theory. That is to say, in examples such as (20) and (21), where there are inaccessible antecedents in the relative clauses that appear in between brackets, the participants in the study tended to compute *Molly* as a possible antecedent of *drove herself* because it matches in gender. However, this was not the case with Darrin, a mismatched inaccessible antecedent (Runner & Head 2014: 8):

(20) The pharmacist(f) [that Molly met] drove herself to the party.

(21) The pharmacist(f) [that Darrin met] drove herself to the party.

Storoshenko (2017), on the other hand, studied the interpretation made by L1 speakers of English ($n = 40$) of reflexives in three different contexts (object *self*, by-phrase *self* and adverbial *self*, as shown in (22a–c), respectively) and *one*-anaphora (as in (22d) in cases of VPE with gender (mis)matches. In his work, participants had to perform a contextual felicity task wherein target sentences with VPE were presented together with a short text giving background information about the situation. The subjects taking part in the study were then asked to rate on a 7-point Likert scale the degree to which the target matched the context using sentences such as the ones in (22), which contain gender mismatch conditions:

(22) a. Josh burned himself, and Peg did too.
b. Tom found a new job by himself, and Claire did too.
c. Hannah is afraid of spiders herself, and Arnold is too.
d. Mae wanted an expensive portrait of her kids, and Alan wanted a cheaper one.

Interestingly, Storoshenko (2017) found that, among the three reflexive cases explored, participants favoured the sloppy interpretation and that there were no gender mismatch effects. More precisely, in the case of object *self* sentences, which are the ones that resemble the experimental items in our study, the results showed that in both gender match and mismatch conditions, native speakers of English significantly ($p < 0.001$) preferred the sloppy reading (6.09 for both match and mismatch) over the strict one (3.71 for match and 4.09 for mismatch). However, in the case of *one*-anaphora constructions, his findings suggested that native speakers of English do not make a strict/sloppy distinction, except for a slight drop on the sloppy gender mismatch condition (which he attributed to a reduced number of participants systematically rejecting those target sentences, that is to say, assigning ratings of 1–2 while other participants assigned 6–7).

2.4 L2 acquisition of reflexives and pronoun gender agreement

The last section of the literature review will broach L2 studies on the acquisition of reflexive pronouns, pronoun gender agreement and their interplay with learners' proficiency.

It is well-known that gender mistakes in L2 English pronoun production are very frequent. Antón-Méndez (2010), for instance, observed that Spanish speakers exhibited significantly more *he/she* confusions than any other type of pronoun errors (e.g., number errors, person errors, omission errors). Gender agreement seems to be especially problematic when English learners need to decide the type of possessive determiner that goes with the following noun, as in (23), where we have "The mother", which is feminine gender, and then, following, "boy", which is masculine. Thus, a common mistake would be to say "his boy" when the learner would mean "her boy".

(23) The mother dressed her boy.

In addition, some works have shown positive (Apaloo & Cardoso 2022 for Brazilian) and negative (Muñoz Lahoz 1994 for Spanish; White et al. 2007 for French and Catalan/Spanish; Imaz-Aguirre 2015 for Basque/Spanish) L1 transfer effects. What these studies have found is that pronoun gender agreement seems to improve with cross- linguistic awareness for positive transfer, and with target language proficiency for negative transfer.

Also, the findings of L2 studies on gender agreement in L2 reflexives (see, for instance, Felser & Cunnings 2012 for German; Liang et al. 2018 for Chinese) have shown that, when assigning gender, native speakers and learners behave differently because the former have no problems in this regard since they follow principle A of the Binding Theory. However, in the case of L2 learners, their interpretations show that they have not acquired Principle A of the Binding Theory and do not bind the reflexive locally. This appears to be so because they prefer to resort to discourse-based coreference assignment. This is illustrated in example (24) from Felser & Cunnings (2012), where L1 German learners of English tend to bind the pronoun *she* (underlined) with the reflexive pronoun *herself*.

(24) Helen has worked at the army hospital for years. <u>She</u> noticed that the soldier had wounded himself/herself while on duty in the Far East.

In this regard, it would be worth mentioning for the purposes of this paper that, as has been shown in the literature (Thomas 1989; 1991; 1993; Padilla, 1990; Eguren 2013), both Spanish and English languages bind reflexive pronouns locally, unlike other languages such as Chinese or Japanese, where both long-distance and local binding would be possible. Therefore, one may assume that Spanish speakers' knowledge about English binding conditions would be similar to English native speakers'. Paradoxically, Thomas (1989) found no statistically significant differences between Chinese- and Spanish-speaking groups. What is more, Chinese learners exhibited a more similar performance to English native speakers as compared to Spanish learners, as the Chinese group bound 69% of the reflexives locally, whereas the Spanish one produced 60% of local responses. In other words, Spanish learners of English did bind the reflexives long-distance even if it was not expected either in their L1 or L2.

There have also been some L2 studies that have analysed the effect that target language proficiency may exert on the acquisition of pronouns and ellipsis so as to check whether more advanced learners' behaviour resembles that of native speakers more than less proficient learners' behaviour. Research has revealed that proficiency has an impact on the acquisition of various aspects of L2 pronouns, for instance, cliticisation (Scuitti 2020), gender agreement (Dong et al. 2015) and anaphora resolution (Contemori et al. 2019). Interestingly, concerning reflexive pronouns, the effect of this variable has shown that the higher the L2 learners' proficiency, the better their acquisition of reflexives (Yip & Tang 1998). In other words, higher proficient learners are capable of treating the binding properties of the L2 reflexives as an independent system from their L1.

Finally, with regard to the literature that has broached the interplay between proficiency, ellipsis and the interpretation of reflexives, research has revealed that higher-proficiency learners' readings resemble more those by native speakers than lower-proficiency learners' ones (Ying 2005; Epoge 2012; Park 2016; Gallardo-del-Puerto & Gandón-Chapela 2024). As will be briefly described below, Ying (2005), Park (2016) and Gallardo-del-Puerto & Gandón-Chapela (2024) investigated L2 English learners' readings of reflexives in bare, referential, and non-referential contexts (instantiated in examples (25)–(27), respectively), whereas Epoge (2012) studied only the contrast between bare and referential contexts.

(25) John defended himself and Bill did too.

(26) John defended himself and Bill did too. Bill was a good friend of John.

(27) John defended himself and Bill did too. Bill went to the restaurant afterwards.

(Gandón-Chapela & Gallardo-del-Puerto 2019: 79)

Crucially, examples (25) and (27) are ambiguous as to whether VPE would be interpreted strictly (*Bill defended John*) or sloppily (*Bill defended Bill*). However, in the case of (26), the contextual information offered biases the reading towards a strict interpretation.

In Ying's (2005) study the participants were 50 L1 Chinese university learners of English (intermediate (n = 28), advanced (n = 22)) studying English majors in Shangai and a control group of 20 native speakers of American English. The findings of this work show that in bare contexts all participants, including the native speakers, interpreted the VPE sloppily (ranging from 74.1% to 82.1%). However, intermediate learners were the ones who did so to a larger extent, the differences attested across the two learner groups and between the intermediate learners and the control group being statistically significant. Nevertheless, no statistically significant differences were found between the advanced learners' group and the control one. Concerning referential contexts, both intermediate and advanced learners favoured the strict interpretation, and so did the native speakers (from 64.9% to 82.1%). This time the differences

found among the three participant groups were statistically significant. Finally, in the case of non-referential contexts, neither the learner groups nor the native speakers showed a marked preference for either of the two possible readings, as no statistically significant differences were attested (from 52.5% to 55.8% for the sloppy reading).

Park (2016) replicated Ying's (2005) study with 26 fourth-year university students in Busan (South Korea) (intermediate (n = 13), advanced (n = 13)) and a control group of 10 native speakers of English. As far as bare contexts are concerned, intermediate and advanced learners of English favoured the sloppy interpretation similarly (91.7% and 88.5%), which was significantly different from that of natives, as they did so to a lesser extent (79.2%). Regarding referential contexts, the control group showed a slight preference for the strict reading (57.5%), the intermediate learners showed a clear preference for the sloppy one (64.1%), and advanced learners chose either interpretation similarly (sloppy 49.4% vs. strict 50.6%). In this case, the interpretations made by intermediate learners were significantly different compared to the two other groups, since no differences were found between the advanced and the control groups. Finally, in the case of non-referential contexts, the sloppy interpretation was the preferred one by all of the groups (ranging from 78.2% to 94.2%), especially by native speakers and advanced learners. One more time, no statistical differences were found between the control and advanced groups, while the intermediate one significantly differed from them.

Gallardo-del-Puerto & Gandón-Chapela (2024), on their part, provided data from 104 Spanish learners of English (A2 (n = 32), B1 (n = 37), and B2 (n = 31) levels according to the Common European Framework of Reference for Languages) from the University of Cantabria (Spain) and a control group of 32 native speakers of British or American English. Their results revealed that participants preferred the sloppy interpretation of reflexives in bare (ranging from 66.4% to 84.4%) and nonreferential (from 58% to 81.2%) contexts, while strict readings were predominant in referential ones (from 70% to 92.2%). They also reported significant differences in the interpretation of learners when compared to native speakers, the differences among the three learner groups studied not being so marked. However, the group with the lowest proficiency level distanced itself most from native speakers. These findings partially confirmed previous research and the authors attributed the discrepancies attested to extraneous variables such as the learners' L1, the range of the proficiency levels under scrutiny, or the characteristics of the control groups that took part in the studies.

Lastly, Epoge (2012) conducted a similar study with 128 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). In this case, no control group of native speakers was included. His work focused only on two of the contexts mentioned so far: bare and referential. Concerning bare contexts, the higher the proficiency level, the more favoured the sloppy interpretation was (ranging from 53.1% to 68.6%). On the other hand, in referential

contexts the opposite tendency was attested since strict interpretations increased as the level of proficiency raised (from 78.3% to 91.9%). It should be noted, however, that in this work no inferential statistics was performed on the data.

What can be gathered from the L2 literature review offered is that one variable that has been unexplored is the effect of gender (mis)match on the readings of English reflexive pronouns (as either strict or sloppy) in cases of VPE (see Storoshenko 2017 for L1 speakers of English). Additionally, the fact that Spanish and English bind the reflexives locally (Thomas 1989; 1991; 1993; Padilla, 1990; Eguren 2013), but Spanish learners of English have been found to disregard their L1 knowledge and sometimes bind the reflexives long-distance (Thomas 1989) would justify a further look into both English native speakers and Spanish learners of English.

3 Research questions

In the present paper, three research questions have been considered in order to explore the effect of gender (mis)match on the interpretation of English reflexive pronouns in cases of VPE by native speakers of English and L1 Spanish learners of English. On the one hand, the first research question (RQ1) addresses the issue of whether gender (mis)match has an impact on participants' readings of English reflexive anaphora in VPE. Research question 2 (RQ2), on the other hand, tackles the issue of whether the direction of the gender mismatch (feminine antecedent-masculine target of ellipsis and vice versa) has an impact on participants' readings of English reflexive anaphora in VPE. Finally, research question 3 (RQ3) examines whether target language proficiency influences participants' readings of English reflexive anaphora in VPE.

RQ1. Does gender mis(match) affect native speakers' and learners' interpretation of English reflexive anaphora in VPE?

RQ2. Does the direction of the gender mismatch affect native speakers' and learners' interpretation of English reflexive anaphora in VPE?

RQ3. Does English proficiency affect participants' interpretation of English reflexive anaphora in VPE?

4 Method

4.1 Participants

Fifty-two Spanish-speaking University of Cantabria (Spain) students learning English and 32 native speakers (NS henceforth) of British or American English participated in this study (22 females and 10 males, average age 38.9, and age range 19–71). The non-native participants were students that were taking BA degrees in Early Childhood and Primary Education Teacher Training, as well as MA students of Second Language Learning and Teaching. As in previous

research (see Ying 2005; Xiao 2015; Gallardo-del-Puerto & Gandón-Chapela 2024), a placement test was used in order to check the learners' proficiency levels, i.e., the Quick Placement Test (Oxford University Press). Traditionally, there have been four types of proficiency determinants in the field of applied linguistics: standardised tests scores, course or grade levels, length of formal instruction, and duration of residence in the target language community. However, according to Xiao (2015), since standardised test scores allow for comparison across test takers, they are among the most reliable indicators of proficiency. Therefore, in line with this view, we have decided to use the Quick Placement Text (OUP) in its forty-item version, which measures students' proficiency up to C1 level of the Common European Framework of Reference for Languages (CEFR). The results of this English proficiency test indicated that our participants' level of English ranged between B1 and C1 levels. According to the Oxford Placement Test (OPT henceforth) assessing guidelines, a score of 24–30 indicates a B1 level, one of 31–35 a B2 level and one of 36–40 a C1 level. Three different proficiency groups were attested among the participants of this study. The first group was formed by 19 B1-level students (18 females and 1 male) whose average age was 22.21 (age range: 20–28) and OPT mean score was 28.15 (24–30). The second group was formed by 22 B2-level students (14 females and 8 males) whose average age was 22.9 (age range: 20–44) and OPT mean score was 32.59 (31–35). Lastly, there was another group composed by 11 C1-level students (9 females and 2 males) whose average age was 22.45 (age range: 20–29) and OPT mean score was 37.81 (36–40). These data have been gathered in **Table 1** for visual convenience:

Proficiency level	Gender	Mean Age	OPT mean score
B1	18 Female 1 Male	22.21 (20–28)	28.15 (24–30)
B2	14 Female 8 Male	22.9 (20–44)	32.59 (30–35)
C1	9 Female 2 Male	22.45 (20–29)	37.81 (36–40)
NS	22 Female 10 Male	38.9 (19–71)	

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

4.2 Instruments

In this study we have used two kinds of instruments: an English proficiency test measuring one of the independent variables of the study, that is, target-language proficiency (as reported in section 4.1); and one two-alternative (sloppy-strict) forced-choice task to check the impact of

gender (mis)match on participants' interpretations of English reflexive anaphora in cases of VPE. This task contained 10 gender-matched and 10 gender-mismatched sentences as in (28) and (29), and 20 distractors. The direction of the gender mismatch (feminine antecedent-masculine target of ellipsis and vice versa) was also considered, since the gender-mismatched experimental items contained half feminine-masculine and half masculine-feminine directions. Examples (28) and (29)–(30) illustrate gender match and mismatch between the reflexive pronoun that acts as the antecedent and that of the ellipsis site, respectively. As for the direction of the gender mismatches, example (29) displays a masculine-feminine ordering, and (30) represents a feminine-masculine one. It must be noted that the gender (mis)match was indicated only by using gendered names which were known by the learners, as they were proper names of high frequency. As can be observed, the participants had to choose one of the two options given immediately below:

(28) Peter has voted for himself and Robert has too.
  (i) Robert has voted for Peter.
  (ii) Robert has voted for Robert.

(29) Peter has voted for himself and Elaine has too.
  (i) Elaine has voted for Peter.
  (ii) Elaine has voted for Elaine.

(30) Elaine has voted for herself and Peter has too.
  (i) Peter has voted for Elaine.
  (ii) Peter has voted for Peter.

4.3 Data Gathering Procedure

The data were gathered in one session during the students' regular class time. Firstly, students had to take the Quick Placement Test (OUP), which would measure their English language proficiency level. They responded to the 40 multiple-choice questions in this test in around 30 minutes. Then, 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) and the two-alternative forced-choice task via Google forms (using either their phones or laptops), which they completed in around 10 minutes on average. In the case of the native speakers, they also completed the short biographical and linguistic survey together with the forced-choice task, but were exempt from taking the Quick Placement Test. As in previous research (Ying 2005; Gandón-Chapela & Gallardo-del-Puerto 2019; Gallardo-del-Puerto & Gandón-Chapela 2024), in the forced-choice task participants were asked to choose one of the two interpretations provided which matched their initial interpretation of the underlined part of the sentence. That is to say, the participants were asked to select the answer that came to their minds first, i.e., following their initial intuition and without going back or making any changes after their initial choice.

4.4 Data analysis procedure

Data were quantitatively analysed with the help of the *Statistical Package for Social Sciences* (SPSS). Concerning descriptive statistics, interpretation preferences (mean scores) and standard deviations were calculated. These figures informed about the central tendency and the dispersion of the data, respectively, as regards the interpretations of English reflexive anaphora in cases of VPE in each of the four research groups (B1, B2, C1 and NS). These figures were calculated for each of the two (mis)matched contexts separately –10 experimental sentences with gender matches and 10 experimental sentences with gender mismatches. Similarly, descriptive statistics are provided for the two gender orderings in mismatched contexts separately –5 experimental sentences with masculine-feminine orderings and 5 experimental sentences with feminine-masculine ones.

As for inferential statistics, the first analyses performed were those pertaining to normality. In that sense, Kolmogorov-Smirnov tests distinguished between those cases where parametric procedures were to be used (since a normal distribution of the samples was found), and those where non-parametric procedures should be employed (as there was at least a sample whose distribution was skewed). (Parametric) T-tests or (non- parametric) Wilcoxon signed-rank tests were alternatively computed to explore differences between the two contexts appearing in the forced-choice task sentences, i.e., gender match vs. gender mismatch, as well as between the two gender orderings, i.e., masculine-feminine vs. feminine-masculine. As for the inquiry into proficiency, (non-parametric) Kruskal Wallis one-way analyses of variance were employed to look into comparisons among the four research groups. Whenever significant differences were discovered, (non-parametric) post-hoc Mann-Whitney tests established all pertaining pairwise comparisons (B1-B2, B1-C1, B1-NS, B2-C1, B2-NS, C1-NS). Statistical probability was considered significant when alpha reached the levels of 0.05(*), 0.01(**) and 0.001(***), respectively. In addition, effect sizes were calculated, with Cohen's d values above 0.2, 0.5 and 0.8 being considered of a small, medium and large strength of difference, respectively.

5 Results

This section will offer a look at statistical results concerning the gender match vs. mismatch contrast, as well as the direction of gender orderings in mismatched contexts. Accordingly, **Table 2** displays the means and standard deviations of the participants' sloppy vs. strict interpretations of VPE reflexives in both gender-matched (GM) and gender-mismatched (GMM) contexts. **Table 3** presents the inferential procedures that will help to discover whether the gender (mis)match variable exerts any significant influence on participants' interpretations of English reflexive anaphora in VPE, whilst **Table 4** focuses on the statistics computed to probe any significant language proficiency effects in said interpretations. **Table 5** shows the means and standard deviations of the participants' sloppy vs. strict interpretations of VPE reflexives in both masculine-feminine (MF) and feminine-masculine (FM) mismatched contexts. Finally, **Table 6** provides the inferential statistics results regarding the direction of the gender mismatch.

Means (Standard deviations)	GM sloppy	GM strict	GMM sloppy	GMM strict
B1 (n = 19)	6.05 (2.87)	3.95 (2.87)	6.26 (2.42)	3.74 (2.42)
B2 (n = 22)	6.86 (2.78)	3.14 (2.78)	7.23 (2.22)	2.77 (2.22)
C1 (n = 11)	7.64 (2.25)	2.36 (2.25)	7.55 (2.06)	2.45 (2.06)
NS (n = 32)	8.75 (1.83)	1.25 (1.83)	7.03 (2.38)	2.97 (2.38)

Table 2: Interpretation preferences (and standard deviations) of VPE reflexive anaphora sloppy/strict interpretations in gender-matched (GM) and gender-mismatched (GMM) contexts.

Descriptive statistics (**Table 2**) showed that all research groups chose sloppy interpretations over strict ones in both gender-matched and mismatched contexts. However, as regards intragroup comparisons between the two contexts, B1 and B2 learners exhibited a slightly stronger preference for sloppy interpretations in gender- mismatched than in gender-matched contexts, whereas C1 and NS groups displayed an opposite trend, that is, they interpreted sentences sloppily comparatively more often in gender-matched than in gender-mismatched contexts. As for intergroup comparisons, the data from gender-matched sentences revealed that the higher the proficiency, the more striking the choice of sloppy readings. This choice ranged from a mean of 8.75 out of 10 in the case of the most proficient speakers (NS) and one of 6.05 in the case of the least (B1). With regard to gender mismatched contexts, this tendency was observed if we look at the data from non-native speakers, with least proficient learners (B1) showing the weakest preference for sloppy readings (6.26) and the most proficient learners exhibiting the strongest choice (7.55). However, contrary to what happened in gender-matched contexts, NSs were not the ones who interpreted sentences sloppily the most in gender-mismatched contexts (7.03).

As for inferential statistics, normality tests indicated that all samples were non- normally distributed with the exception of the data provided by C1 learners. Pertaining intragroup comparisons between gender-matched and gender-mismatched contexts (see **Table 3**), they turned out to be statistically significant only in the case of the NS group, as indicated by the outcome of the Wilcoxon signed-rank test. Besides, the magnitude of the effect size was large. In

GM vs. GMM comparisons	Wilcoxon		t-test		Cohens' d
	z	p	t	p	
B1 (n = 19)	–.608	.543			–.079
B2 (n = 22)	–.578	.563			–.147
C1 (n = 11)			–.161	.875	.042
NS (n = 32)	–4.186	000**			.810

Table 3: Comparisons between sloppy/strict interpretations of VPE reflexive anaphora in gender-matched (GM) vs. gender-mismatched (GMM) contexts in each participant group.

other words, NSs interpreted the English reflexive pronoun sloppily significantly more often in the sentences where there was a gender match than in those with a gender mismatch. However, for all learner groups, no statistical significance was reached for their slightly stronger choice of sloppy readings.

Regarding intergroup comparisons (**Table 4**), the data coming from gender-matched contexts revealed that there were significant differences among the four research groups, as indicated by the Kruskal Wallis one-way analysis of variance (H = 14.735, chi-square = 14.492, mean = 8, p = 0.02*). Post-hoc pairwise comparisons through Mann-Whitney tests indicated that significant differences with a large effect size were discovered when the NS group was compared to the B1 and B2 learners, but not when compared to the C1 learners. In other words, NSs interpreted gender-matched sentences sloppily to a larger extent than intermediate proficiency learners, whereas advanced learners' choice was similar to that of natives in such a context, even though statistical significance could be masked by the lower number of participants in the advanced group. As regards comparisons among learners themselves, no statistically significant differences were attested as to their gender-matched sentence interpretations.

Proficiency effect (intergroup comparisons)	**GM**			**Cohens' d**
	Mann-Whitney U	**z**	**p**	
B1 (n = 19) – **B2** (n = 22)	176.000	–.872	.383	–.287
B1 (n = 19) – **C1** (n = 11)	71.000	–1.454	.158	–.617
B1 (n = 19) – **NS** (n = 32)	122.000	–3.657	.000**	–1.122
B2 (n = 22) – **C1** (n = 11)	102.000	–.739	.486	.308
B2 (n = 22) – **NS** (n = 32)	216.500	–2.505	.012*	–.803
C1 (n = 11) – **NS** (n = 32)	176.000	–1.709	.104	–.541

Table 4: Pairwise comparisons between participant groups on the choice of VPE reflexive anaphora sloppy/strict interpretations in gender-matched contexts.

In relation to participants' readings of gender-mismatched sentences, the analysis of variance revealed no statistically significant differences among the four research groups (H = 2.414, chi-square = 3.925, mean = 7, p = .491). That is to say, all participant groups chose the sloppy interpretation over the strict one in gender-mismatched sentences to a similar degree.

As for the descriptive statistics of the results regarding the direction of the gender mismatch (**Table 5**), what can be gathered is that the four research groups obtained higher sloppy interpretation means (max = 5) than strict interpretation means in both gender-mismatched contexts, even though this preference was stronger for all groups when the direction of the gender mismatch was feminine-masculine than when it was in masculine-feminine. As for

Means (Standard deviations)	MF sloppy	MF strict	FM sloppy	FM strict
B1 (n = 19)	2.68 (1.53)	2.32 (1.52)	3.58 (1.22)	1.42 (2.42)
B2 (n = 22)	3.23 (1.54)	1.77 (1.54)	3.95 (1.00)	1.05 (1.00)
C1 (n = 11)	3.36 (1.36)	1.64 (1.36)	4.36 (.81)	0.64 (.81)
NS (n = 32)	3.22 (1.78)	1.78 (1.78)	3.75 (.92)	1.25 (.92)

Table 5: Interpretation preferences (and standard deviations) of VPE reflexive anaphora sloppy/strict interpretations in gender-mismatched masculine-feminine (MF) and feminine-masculine (FM) contexts

intergroup differences, it is in C1 learners where the sloppy reading was more striking in both mismatched contexts (masculine-feminine: 3.36; feminine-masculine: 4.36), even more than in NSs, whereas B1 learners were the ones who chose the sloppy interpretation the least in both gender-mismatched contexts (masculine-feminine: 2.68; feminine-masculine: 3.58).

Inferential analyses comparing the two mismatched contexts (**Table 6**) revealed that the choice of sloppy readings, as compared to the strict ones, was significantly more abundant in feminine-masculine sentences than in masculine-feminine ones in the four participant groups. In other words, there was a significant effect of the direction of the gender mismatch because when the sentence containing the antecedent had a female subject and that of the target of ellipsis a male subject, the probabilities of a sloppy reading increased in a significant manner, as compared to those sentences with male subjects as antecedents and female subjects in the target of ellipsis. This behaviour, besides, occurred all across the board regardless of participants' proficiency in English.

MF vs. FM comparisons	Wilcoxon		Cohens' d
	z	p	
B1 (n = 19)	–2.494	.013*	.650
B2 (n = 22)	–2.429	.015*	.555
C1 (n = 11)	–2.456	.014*	.893
NS (n = 32)	–1.972	.049*	.374

Table 6: Comparisons between sloppy/strict interpretations of VPE reflexive anaphora in gender-mismatched masculine-feminine (MF) and feminine-masculine (FM) orderings in each participant group.

As for intergroup comparisons of ellipsis interpretations, the Kruskal-Wallis test revealed that differences among the four participant groups did not reach statistical significance in either feminine-masculine (H = 2.230; p = .551) or masculine-feminine (H = 4.626; p = .201) orderings. This finding replicates the tendency for all mismatched contexts described above (see

Table 2), which indicates that language proficiency exerted no effect on the interpretations of gender-mismatched sentences overall.

6 Discussion

This section discusses the results mentioned above according to the three research questions of the present paper, which aim at determining the effects of both gender (mis)match and the direction of the gender mismatch on the interpretation of English reflexive pronouns in cases of VPE. Additionally, we investigate the effect of English proficiency on participants' readings of English reflexive anaphora in cases of VPE in the above-mentioned contexts.

Concerning the first research question (*Does gender mis(match) affect native speakers' and learners' interpretation of English reflexive anaphora in VPE?*), our results show that the sloppy reading was always preferred, regardless of gender (mis)match. This would be in line with the findings reported in Storoshenko (2017), whose native speakers of English assigned a higher score to the VPE sentences where the written scenario previously provided biased a sloppy interpretation of the reflexive pronoun, as compared to those contexts biasing a strict reading. The data have also confirmed that L1 speakers –but not Spanish L2 English learners– attribute a sloppy interpretation to gender-mismatched configurations significantly less often than they do to gender-matched ones. Thus, it could be affirmed that gender mismatch makes strict readings (that is, non-local, long-distance binding) more probable in the interpretation of reflexives in L1 English VPE (evincing that Principle A of the Binding Theory is challenged). It would be as if gender mismatch creates an interference in the strong tendency in L1 English for reflexive pronouns to be locally bound in VPE (unless pragmatic cues indicate otherwise; see Ying 2005; Park 2016; Gandón-Chapela & Gallardo-del-Puerto 2019). This outcome would be contrary to the preliminary findings of Storoshenko's (2017) study, where there was no effect of gender (mis)match in object *self* constructions. However, our results are in line with Runner & Head's (2014) finding in their eye-tracking experiments that the interpretation of reflexive pronouns is sensitive to gender (mis)match effects. It would also agree with Aparicio et al.'s (2015) effects of gender mismatch on stripping in Spanish. However, while Aparicio et al. (2015) would regard their findings as evidence for a structural approach to ellipsis, inasmuch as the interpretation of inflectional features is relevant for ellipsis computation (contra Abeillé & Kim 2022 for French stripping, who argue for a non-structural account of their data), the particular case of reflexive anaphora in VPE may not be so straight forward. One should bear in mind that, in the strict reading, there would be no syntactic identity, even in cases with gender match (*Sienna$_i$ criticised herself$_i$ and Melanie did ~~criticise her~~$_i$ too*). Therefore, the choice between strict and sloppy interpretations would imply choosing between semantic and syntactic identity (*Sienna$_i$ criticised herself$_i$ and Melanie$_j$ did ~~criticise herself~~$_j$ too*). Concerning those sentences with gender mismatch, there would be reflexive mismatch in the strict reading (*Sienna$_i$ criticised herself$_i$ and George$_j$*

did ~~criticise her~~$_i$ too) and gender mismatch in the sloppy reading (*Sienna$_i$ criticised herself$_i$ and George$_j$ did ~~criticise himself~~$_j$ too*). Therefore, in this type of sentences there would be no syntactic identity. All this may be the reason why the strict reading is preferred in those sentences with gender mismatch, given that the sloppy interpretation would be more costly to process than those sentences with gender match. Another issue that merits attention would be the fact that the results attested may be due to overall processing constraints on coordination, which seem to favour parallelism independently of ellipsis (see Frazier et al. 1984). Finally, an alternative account of the impact of gender (mis)match on the interpretation of English reflexives in VPE would be related to its discourse conditions, as VPE may be either auxiliary-focus or subject-focus. The literature on the issue has revealed that the auxiliary-focus condition licenses more mismatches (see Kertz 2013; Miller & Pullum 2014). Therefore, as already mentioned in section 2.1, since the experimental sentences in the present study are all subject-focus, gender (mis) match might also be playing a role, at least for this group of native speakers of English. To sum up, our findings on the interpretation of English reflexives in cases of VPE with gender mismatch would provide supportive evidence in favour of a non-structural approach to ellipsis in which 'what you see is what you get'.

As for Spanish learners of English, the absence of a gender (mis)match effect in their data would allow one to speculate that L2 learners do not suffer from such effect or interference, as their preference for sloppy readings was already less strong than that of native speakers overall (probably as a consequence of their non-nativeness, see Gandón- Chapela & Gallardo-del Puerto 2019), that is, because the sloppy (local, short-distance binding) reading is not so marked in L2 English (see Thomas 1989). This finding is in line with research on gender agreement in L2 reflexives (see Felser & Cunnings 2012; Liang et al. 2018) showing a differential behaviour by native and non-native speakers when assigning gender. This appears to be so because native speakers follow principle A of the Binding Theory and thus bind the reflexive locally, whereas L2 learners do not tend to do so, since they show a preference for more salient (topic or first-mention) referents.

Regarding the second research question (*Does the direction of the gender mismatch affect native speakers' and learners' interpretation of English reflexive anaphora in VPE?*), a significant effect of gender orderings (masculine-feminine vs. feminine-masculine) was discovered for all the participant groups. More precisely, participants more strikingly opted for sloppy readings of reflexives in cases of VPE when the sentence containing the antecedent had a female subject and that of the target of ellipsis a male one (e.g., *Elaine has voted for herself and Peter has too*), as compared to those sentences with male subjects as antecedents and female subjects in the target of ellipsis (e.g., *Peter has voted for himself and Elaine has too*). Our results are in line with previous findings reported by Aparicio et al. (2015), who found that the direction of gender orderings in mismatched Spanish stripping constructions played a role in their acceptability, since the

ordering masculine-feminine was judged more acceptable than the ordering feminine-masculine. Sprouse et al. (2022) also found the same results for cases of nominal ellipsis, where the feminine antecedent (*Mary is an actress and John is too*) was judged less acceptable than the masculine one (*John is an actor and Mary is too*). The present study, where the data attested a preference for sloppy readings in feminine-masculine orderings, would follow the same tendency: when participants were presented with a feminine antecedent, they more strikingly bound the reflexive locally in the ellipsis site and thus, the number of strict readings decreased. This means that they tend to disprefer a feminine pronoun in the ellipsis site (which would give rise to a strict interpretation), as it may be regarded as more unacceptable than its masculine counterpart. This seems to point to the fact that participants would be more inclined to consider the masculine pronoun as the default one.

As far as the third research question is concerned (*Does English proficiency affect participants' interpretation of English reflexive anaphora in VPE?*), our data show that a proficiency effect was attested in the case of gender-matched sentences, given that native speakers' sloppy choice was statistically superior to that of B1 and B2 learners. This result would be in line with the different behaviour attested in native and non-native speakers when assigning gender to L2 reflexives, as NSs have no problems following Principle A of the Binding Theory, whereas non-natives tend to resort to discourse reference assignment (Felser & Cunnings 2012; Liang et al. 2018). Our findings would also agree with the effect of proficiency discovered in Ying's (2005) study on the interpretations of L2 reflexives in VPE, since, as in the present paper, he found a significantly different behaviour between NSs and intermediate learners, but not between NSs and advanced learners, contrary to Park (2016), who found statistical differences between advanced learners and NSs. However, our results replicate Gallardo-del-Puerto & Gandón-Chapela's (2024) study with Spanish learners of English in that intermediate learners chose the sloppy interpretation significantly less than NSs. These divergent results among the different studies mentioned above, as has been stated in the literature review, could be due to differences in the instruments used to measure proficiency in the different studies. On the other hand, notice that no proficiency effect was discovered in the mismatched experimental sentences, as NSs lowered the number of sloppy interpretations and, thus, their readings were more similar to non-native speakers'. The lack of proficiency effects in mismatched contexts was further confirmed when the direction of the gender mismatch was explored, since no statistical differences were attested in either masculine-feminine or feminine-masculine orderings.

7 Conclusions and issues for further research

This paper aimed to explore the effect of gender (mis)match and the direction of the gender mismatch on the interpretation of English reflexive pronouns in cases of VPE by native speakers of English and L1 Spanish learners of English. As has been pointed out in the discussion section,

our findings have revealed two main facts as far as NSs of English are concerned: (i) they always favour the sloppy reading and (ii) they are sensitive to gender mismatches when interpreting reflexives in VPE constructions. A close analysis of our data seems to support a non-structural account of VPE resolution. As for learners, they also tended to prefer the sloppy reading. However, no gender mismatch effects were discovered, given that their preference for sloppy readings was less marked overall in both matched and mismatched contexts. Additionally, a proficiency effect was found in matched sentences, whereby intermediate learners interpreted the experimental sentences sloppily significantly less frequently than NSs, whereas advanced learners did not. Finally, both NSs and L2 Spanish learners of English coincided in exhibiting the same effect of the direction of gender ordering, namely, there was a more marked preference for sloppy readings in feminine-masculine orderings.

In future stages, it would be convenient to conduct an additional independent task, such as a truth-value judgement task (Crain & Thornton 1998) or think-aloud protocol (Yoshida 2008) that could help to shed light on why participants chose one reading over another. As in previous studies (Felser & Cunnings 2012; Runner & Head 2014; Aparicio et al. 2015), it would also be crucial to measure online processing of the effect of gender (mis)match on the interpretation of English reflexive pronouns in cases of VPE by carrying out eye-tracking experiments that would offer the possibility of gathering data from participants' reaction times.

It should also be noted that this paper has been, to our knowledge, the first to focus on the effect of gender (mis)matches on the interpretation of English reflexives in cases of VPE by L1 Spanish learners of English. Complementarily, since in Spanish the reflexive can be gender neutral (*defenderse*) or not (*defenderse a sí mismo/misma*) a look into these learners' behaviour with this type of constructions in their L1 would shed light on the effect of gender (mis)match.

In further research, it would be of vital importance to analyse the variables that have already been tackled in L1 studies, such as the impact that discourse connectives (Hestvik 1995; Kehler 2000; Frazier & Clifton 2006; Kim & Runner 2009; Ong & Brasoveanu 2014; McKillen 2016), verb semantics (Dalrymple et al. 1991; Ong & Brasoveanu 2014), syntactic configuration (Kim & Runner 2009; Hestvik 1995; Frazier & Clifton 2006; McKillen 2016), or negation (Ong & Brasoveanu 2014) may have on the readings of English reflexive pronouns in cases of VPE.

Finally, it would be crucial to add data from Advanced C2 level L1 Spanish learners of English, as well as analysing the effect of participants' gender (by increasing the number of males in the sample), given that it could also play a role in VPE resolution. Accordingly, it would be interesting to investigate whether, for instance, male participants may prefer male denoting stimuli and female participants female ones. To be sure, the study of all these variables will offer the opportunity of exploring new research avenues in the field of SLA that will also help to advance in the theoretical knowledge of the ellipsis domain.

Abbreviations

CEFR: Common European Framework of Reference for Languages

COCA: Corpus of Contemporary American English

[e]: ellipsis

FM: Feminine-Masculine

FEM-SG: Feminine Singular

GM: Gender Match

GMM: Gender Mismatch

IT-NEUT: It Neuter

L1: First Language

L2: Second Language

MF: Masculine-Feminine

MASC-SG: Masculine Singular

NS: Native Speaker

OPT: Oxford Placement Test

PAE: Post-Auxiliary Ellipsis

SLA: Second Language Acquisition

VPE: VP-ellipsis

Data accessibility statement

Gandón-Chapela, Evelyn & Gallardo-del-Puerto, Francisco (2025, March 12). Database on the effect of gender (mis)match on the interpretation of English reflexive pronouns in cases of VP-ellipsis by native and non-native speakers of English. Retrieved from osf.io/wsj3x.

Ethics and consent

Before performing the two-alternative forced-choice task, participants were asked to fill in this consent written in both English and Spanish by selecting either of the two options given below:

I authorise Evelyn Gandón-Chapela and Francisco Gallardo-del-Puerto, professors of the Department of Philology at the University of Cantabria, to make use of the data collected during the academic year 2022/2023 for research purposes only, understanding that the right of confidentiality and anonymity of the personal information contained therein is guaranteed.

Autorizo a los profesores del departamento de Filología de la Universidad de Cantabria Evelyn Gandón-Chapela y Francisco Gallardo-del-Puerto a hacer uso de los datos recogidos en el curso

académico 2022/2023 con fines exclusivamente de investigación, entendiendo que se garantiza el derecho de confidencialidad y anonimato de la información personal contenida en los mismos.

□ I give my consent/ Doy mi consentimiento

□ I do not give my consent/ No doy mi consentimiento

Acknowledgements

We are grateful to the editors and anonymous reviewers, whose comments have helped to enhance the quality of this study. Special thanks also go to Prof. Javier Pérez-Guerra for his generous assistance to further improve this paper.

Competing interests

The authors have no competing interests to declare.

References

Abeillé, Anne & Kim, Jong-Bok. 2022. Me too fragments in English and French: A direct interpretation approach. *The Linguistic Review* 39(3). 495–524. DOI: https://doi.org/10.1515/tlr-2022-2095

Aelbrecht, Lobke. 2009. *You have the right to remain silent. The syntactic licensing of ellipsis*. Brussels: Catholic University of Brussels dissertation.

Aelbrecht, Lobke. 2010. *The syntactic licensing of ellipsis*. Amsterdam: John Benjamins. DOI: https://doi.org/10.1075/la.149

Antón-Méndez, Inés. 2010. Gender bender: Gender errors in L2 pronoun production. *Journal of Psycholinguistic Research* 39(2). 119–139. DOI: https://doi.org/10.1007/s10936-009-9129-z

Apaloo, Marie & Cardoso, Walcir. 2022. Examining the effects of crosslinguistic awareness on the acquisition of English possessive determiners: The case of Brazilian Portuguese speakers. *Language Awareness* 31(3). 307–327. DOI: https://doi.org/10.1080/09658416.2021.1915322

Aparicio, Helena & Franich, Katie & Xiang, Ming. 2015. Morphological feature mismatches under ellipsis: An eye-tracking study. In Bui, Thui & Özyıldız, Deniz (eds.), *NELS 45: Proceedings of the 45th meeting of the North East Linguistic Society: Volume 3*, 1–14. Cambridge, MA: Massachusetts Institute of Technology.

Arregui, Ana & Clifton Jr., Charles & Frazier, Lyn & Moulton, Keir. 2006. Processing elided VPs with flawed antecedents: The recycling hypothesis. *Journal of Memory and Language* 55(2). 232–246. DOI: https://doi.org/10.1016/j.jml.2006.02.005

Bîlbîie, Gabriela. 2017. *Grammaire des constructions elliptiques: Une* étude *comparative des phrases sans verbe en roumain et en français*. (Empirically Oriented Theoretical Morphology and Syntax 2). Berlin: Language Science Press. DOI: https://doi.org/10.5281/zenodo.573753

Bobaljik, Jonathan D. & Zocca, Cynthia Levart. 2011. Gender markedness: The anatomy of a counter-example. *Morphology* 21(2). 141–166. DOI: https://doi.org/10.1007/s11525-010-9156-3

Chomsky, Noam. 1981. *Lectures on government and binding*. The Netherlands: Foris publications.

Chung, Sandra. 2006. Sluicing and the lexicon: The point of no return. In Cover, Rebecca T. & Kim, Yuni (eds.), *Proceedings of the annual meeting of the Berkeley Linguistics Society* 31(BLS), 73–91. Berkeley, CA: Berkeley Linguistics Society. DOI: https://doi.org/10.3765/bls.v31i1.896

Chung, Sandra. 2013. Syntactic identity in sluicing: How much, and why. *Linguistic Inquiry* 44. 1–39. DOI: https://doi.org/10.1162/LING_a_00118

Chung, Sandra & Ladusaw, William & McCloskey, James. 1995. Sluicing and logical form. *Natural Language Semantics* 3. 239–282. DOI: https://doi.org/10.1007/BF01248819

Clifton Jr., Charles & Frazier, Lyn. 2010. Imperfect ellipsis: Antecedents beyond syntax? *Syntax* 13(4). 279–297. DOI: https://doi.org/10.1111/j.1467-9612.2010.00142.x

Contemori, Carla & Asiri, Ohood & Perea Irigoyen, Elva Deida. 2019. Anaphora resolution in L2 English: An analysis of discourse complexity and cross-linguistic interference. *Studies in Second Language Acquisition*, 1–28. DOI: https://doi.org/10.1017/S0272263119000111

Crain, Stephen & Thornton, Rosalind. 1998. The truth value judgment task: Fundamentals of design. *University of Maryland working papers in linguistics* 6. 61–70.

Culicover, Peter & Jackendoff, Ray. 2005. *Simpler syntax*. Oxford: Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780199271092.001.0001

Dalrymple, Mary. 2005. Against reconstruction in ellipsis. In Elugardo, Reinaldo & Stainton, Robert J. (eds.), *Ellipsis and nonsentential speech* (Studies in Linguistics and Philosophy), 31–55. Dordrecht: Springer. DOI: https://doi.org/10.1007/1-4020-2301-4_2

Dalrymple, Mary & Shieber, Stuart M. & Pereira, Fernando C. N. 1991. Ellipsis and higher-order unification. *Linguistics and Philosophy* 14. 399–452. DOI: https://doi.org/10.1007/BF00630923

Davies, Mark. 2008. *The Corpus of Contemporary American English (COCA)*. Available at: https://www.english-corpora.org/coca/

Depiante, Marcela & Masullo, Pascual J. 2001. Género y número en la Elipsis Nominal: Consecuencias para la Hipótesis Lexicalista. Paper presented at *I Encuentro de Gramática Generativa*. Universidad Nacional del Comahue.

Dong, Yanping & Wen, Yun & Zeng, Xiaomeng & Ji, Yifei. 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. DOI: https://doi.org/10.1007/s10936-014-9314-6

Duffield, Nigel G. & Matsuo, Ayumi. 2009. Native-speakers' vs. L2 learners' sensitivity to parallelism in VP-ellipsis. *Studies in Second Language Acquisition* 31. 1–31. DOI: https://doi.org/10.1017/S0272263109090044

Eguren, Luis. 2013. Binding: Deixis, Anaphors, Pronominals. In Hualde, José Ignacio & Olarrea, Antxon & O'Rourke, Erin (eds.), *The handbook Hispanic linguistics*, 556–577. Wiley. https://elibro.net/es/ereader/unican/180838?page=586

Epoge, Napoleon Kang. 2012. Reflexive anaphora in VP-elliptical sentences of ESL learners in Cameroon. *Syllabus Review* 3(1). 257–279.

Felser, Claudia & Cunnings, Ian. 2012. Processing reflexives in a second language: The timing of structural and discourse-level constraints. *Applied Psycholinguistics* 33(3). 571–603. DOI: https://doi.org/10.1017/S0142716411000488

Fiengo, Robert & May, Robert. 1994. *Indices and identity*. Cambridge, MA: MIT Press.

Frazier, Lyn & Clifton Jr., Charles. 2000. On bound variable interpretations: The LF-only hypothesis. *Journal of Psycholinguistic Research* 29. 125–139. DOI: https://doi.org/10.1023/A:1005136826534

Frazier, Lyn & Clifton Jr., Charles. 2005. The syntax-discourse divide: Processing ellipsis. *Syntax* 8(2). 121–174. DOI: https://doi.org/10.1111/j.1467-9612.2005.00077.x

Frazier, Lyn & Clifton Jr., Charles. 2006. Ellipsis and discourse coherence. *Linguistics and Philosophy* 29(3). 315–346. DOI: https://doi.org/10.1007/s10988-006-0002-3

Frazier, Lyn & Taft, Lori & Roeper, Tom & Clifton Jr., Charles & Ehrlich, Kate. 1984. Parallel structure: A source of facilitation in sentence comprehension. *Memory & Cognition* 12(5). 421–30. DOI: https://doi.org/10.3758/BF03198303

Gallardo-del-Puerto, Francisco & Gandón-Chapela, Evelyn. 2024. The interpretation of reflexive pronouns in VP-ellipsis by L2 English learners with different proficiency levels. *Spanish Journal of Applied Linguistics* 37. 424–453. DOI: https://doi.org/10.1075/resla.21054.gal

Gandón-Chapela, Evelyn. 2016. *A corpus-based analysis of Post-Auxiliary Ellipsis in Modern English: Methodological and theoretical issues*. Vigo: University of Vigo dissertation.

Gandón-Chapela, Evelyn. 2020a. *On invisible language in Modern English: A corpus-based approach to ellipsis*. London: Bloomsbury Academic. DOI: https://doi.org/10.5040/9781350064546

Gandón-Chapela, Evelyn. 2020b. A corpus-based analysis of Post-Auxiliary Ellipsis voice mismatches in Late Modern English. Special Issue "Fragments or Ellipsis or both? New Challenges and New Perspectives". *Acta Linguistica Hafniensia. International Journal of Linguistics*, 1–16. DOI: https://doi.org/10.1080/03740463.2020.1765272

Gandón-Chapela, Evelyn & Gallardo-del-Puerto, Francisco. 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. DOI: https://doi.org/10.26754/ojs_misc/mj.20196340

Garnham, Alan & Cain, Jane & Kate, Oakhill. 1998. Selective retention of information about the superficial form of text: Ellipses with antecedents in main and subordinate clauses. *The Quarterly Journal of Experimental Psychology. Section A: Human Experimental Psychology* 51(1). 19–39. DOI: https://doi.org/10.1080/713755747

Ginzburg, Jonathan & Sag, Ivan A. 2000. *Interrogative investigations*. Stanford, CA: CSLI Publications.

Imaz-Aguirre, Ainara. 2015. *The acquisition of gender agreement in L3 English by Basque/Spanish bilinguals*. Vitoria: University of the Basque Country dissertation.

Hankamer, Jorge & Sag, Ivan. 1976. Deep and surface anaphora. *Linguistic Inquiry* 7. 391–428.

Hardt, Daniel. 1993. *Verb phrase ellipsis: Form, meaning and processing*. Philadelphia, PA: University of Pennsylvania dissertation.

Hestvik, Airld. 1995. Reflexives and ellipsis. *Natural Language Semantics* 3. 211–237. DOI: https://doi.org/10.1007/BF01249838

Huddleston, Rodney & Pullum, Geoffrey K. 2002. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/9781316423530

Jacobson, Pauline. 2003. Binding without pronouns (and pronouns without binding). In Kruijff, Geert-Jan M. & Oehrle, Richard T. (eds.), *Resource-Sensitivity, Binding and Anaphora*, 57–96. Dordrecht: Kluwer. DOI: https://doi.org/10.1007/978-94-010-0037-6_3

Jacobson, Pauline. 2008. Direct compositionality and variable-free semantics: The case of antecedent contained deletion. In Johnson, Kyle (ed.), *Topics in Ellipsis*, 30–68. Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511487033.003

Jacobson, Pauline. 2009. Do representations matter or do meanings matter: The case of antecedent containment. In Hinrichs, Erhard & Nerbonne, John (eds.), *Theory and Evidence in Semantics*, 81–107. Stanford, CA: CSLI Publications.

Johnson, Kyle. 2001. What VP ellipsis can do, and what it can't, but not why. In Baltin, Martin & Collins, Chris (eds.), *The handbook of contemporary syntactic theory*, 439–479. Oxford: Blackwell Publishers. DOI: https://doi.org/10.1002/9780470756416.ch14

Keenan, Edward. 1971. Names, quantifiers, and the sloppy identity problem. *Papers in Linguistics* 4. 211–232. DOI: https://doi.org/10.1080/08351817109370257

Kehler, Andrew. 2000. Coherence and the resolution of ellipsis. *Linguistics and Philosophy* 23. 533–575. DOI: https://doi.org/10.1023/A:1005677819813

Kehler, Andrew. 2002. *Coherence, reference and the theory of grammar*. Stanford: CSLI Publications.

Kertz, Laura. 2008. Focus structure and acceptability in verb phrase ellipsis. In Abner, Natasha & Bishop, Jason (eds.), *Proceedings of the 27th West Coast Conference on Formal Linguistics*, 283–291. Somerville, MA: Cascadilla Proceedings Project.

Kertz, Laura. 2010. *Ellipsis reconsidered*. La Jolla, CA: University of California San Diego dissertation.

Kertz, Laura. 2013. Verb phrase ellipsis: The view from information structure. *Language* 89. 390–428. DOI: https://doi.org/10.1353/lan.2013.0051

Kim, Christina & Kobele, Gregory M. & Runner, Jeffrey T. & Hale, John T. 2011. The acceptability cline in VP ellipsis. *Syntax* 14. 318–354. DOI: https://doi.org/10.1111/j.1467-9612.2011.00160.x

Kim, Christina S. & Runner, Jeffrey T. 2009. Strict identity, coherence, and parallelism in VP ellipsis. In Cormany, Ed & Ito, Satoshi & Lutz, David (eds.), *Semantics and Linguistic Theory: Proceedings of the 19th Semantics and Linguistic Theory Conference*, 275–287. DOI: https://doi.org/10.3765/salt.v19i0.2535

Kitagawa, Yoshihisa. 1991. Copying identity. *Natural Language and Linguistic Theory* 9. 497–536. DOI: https://doi.org/10.1007/BF00135356

Kučerová, Ivona & Chapman, Cassandra & Moulton, Keir. 2021. How to value Gender: Lexicon, Agree and Feature Transmission under ellipsis. In Soo, Rachel & Chow, Una Y. & Nederveen, Sander (eds.), *Proceedings of the 38th West Coast Conference on Formal Linguistics*, 253–263. Somerville, MA: Cascadilla Proceedings Project. www.lingref.com, document #3570.

Lasnik, Howard. 1999. Verbal morphology: *Syntactic Structures* meets the Minimalist Program. In Lasnik, Howard (ed.), *Minimalist analysis*, 97–119. Malden, MA: Blackwell.

Liang, Lijuan & Wen, Yun & Dong, Yanping. 2018. Gender constraint in L1 and L2 reflexive pronoun resolution by Chinese-English bilinguals. *Journal of Neurolinguistics* 45. 1–12. DOI: https://doi.org/10.1016/j.jneuroling.2017.08.001

Lobeck, Anne. 1995. Ellipsis. *Functional heads, licensing and identification*. New York & Oxford: Oxford University Press. DOI: https://doi.org/10.1093/oso/9780195091816.001.0001

McKillen, Alanah. 2016. *On the interpretation of reflexive pronouns*. Montreal, Canada: McGill University dissertation. https://www.baclac.gc.ca/eng/services/theses/Pages/item.aspx?idNumber=953107291

Merchant, Jason. 2001. *The syntax of silence: Sluicing, islands, and the theory of ellipsis*. Oxford: Oxford University Press. DOI: https://doi.org/10.1093/oso/9780199243730.001.0001

Merchant, Jason. 2004. Fragments and ellipsis. *Linguistics and Philosophy* 27(6). 661–738. DOI: https://doi.org/10.1007/s10988-005-7378-3

Merchant, Jason. 2008. An asymmetry in voice mismatches in VP-ellipsis and Pseudogapping. *Linguistic Inquiry* 39(1). 169–179. DOI: https://doi.org/10.1162/ling.2008.39.1.169

Merchant, Jason. 2013. Voice and ellipsis. *Linguistic Inquiry* 44(1). 77–108. DOI: https://doi.org/10.1162/LING_a_00120

Merchant, Jason. 2014. Gender mismatches under nominal ellipsis. *Lingua* 151. 9–32. DOI: https://doi.org/10.1016/j.lingua.2014.01.008

Merchant, Jason. 2019. Ellipsis: A survey of analytical approaches. In van Craenenbroeck, Jeroen & Temmerman, Tanja (eds.), *The handbook of ellipsis*, 19–45. Oxford: Oxford University Press. DOI: https://doi.org/10.1093/oxfordhb/9780198712398.013.2

Miller, Philip. 2011. The choice between verbal anaphors in discourse. In Hendrickx, Iris & Devi, Sobha Lalitha & Branco, António & Mitkov, Ruslan (eds.), *Anaphora processing and applications: 8th discourse anaphora and anaphor resolution colloquium (DAARC 2011)*, *Lecture notes in artificial intelligence*, vol. 7099, 82–95. Berlin: Springer.

Miller, Philip. 2014. A corpus study of Pseudogapping and its theoretical consequences. In Piñón, Christopher (ed.), *Empirical issues in syntax and semantics* 10, 73–90. http://www.cssp.cnrs.fr/eiss10/

Miller, Philip & Pullum, Geoffrey K. 2014. Exophoric VP ellipsis. In Hofmeister, Philip & Norcliffe, Elisabeth (eds.), *The core and the periphery: Data-driven perspectives on syntax inspired by Ivan A. Sag*, 5–32. Stanford: CSLI Publications.

Miller, Philip Harold & Hemforth, Barbara. 2014. Verb phrase ellipsis with nominal antecedents. hal-01234401.

Muñoz Lahoz, Carmen. 1994. A case of frequency-based markedness. *Atlantis* XVI(1–2). 165–177.

Nunes, Jairo & Zocca, Cynthia Levart. 2005. Morphological identity in ellipsis. *Leiden Papers in Linguistics* 2. 29–42.

Nunes, Jairo & Zocca, Cynthia Levart. 2009. Lack of morphological identity and ellipsis resolution in Brazilian Portuguese. In Nunes, Jairo (ed.), *Minimalist essays on Brazilian Portuguese syntax*, vol. 142, 215–236. Amsterdam/Philadelphia: John Benjamins Publishing. DOI: https://doi.org/10.1075/la.142.13nun

Ong, Matthew & Brasoveanu, Adrian. 2014. Strict and sloppy reflexives in VP-ellipsis. In Piñón, Christopher (ed.), *Empirical issues in syntax and semantics 10*, 251–268. http://www.cssp.cnrs.fr/eiss10/

Padilla, José A. 1990. *On the definition of binding domains in Spanish*. Dordrecht: Kluwer Academic Publishers. DOI: https://doi.org/10.1007/978-94-009-2033-0

Park, Eugene. 2016. Korean English learners' interpretation of reflexive anaphora in VP- ellipsis. *Journal of Language Sciences* 23(3). 111–135. DOI: https://doi.org/10.14384/kals.2016.23.3.111

Poppels, Till. 2022. Explaining ellipsis without identity. *The Linguistic Review* 39(3). 341–400. DOI: https://doi.org/10.1515/tlr-2022-2091

Poppels, Till & Kehler, Andrew. 2018. Overcoming the identity crisis: Novel evidence for a referential theory of verb phrase ellipsis. In Edmiston, Daniel & Ermolaeva, Marina & Hakgüder, Emre & Lai, Jackie & Montemurro, Kathryn & Rhodes, Brandon & Sankhagowit, Amara & Tabatowski, Michael (eds.), *Proceedings of the annual meeting of the Chicago Linguistics Society,* vol. 53, 291–305. Chicago (IL): Chicago Linguistic Society.

Poppels, Till & Kehler, Andrew. 2019. Reconsidering asymmetries in voice-mismatched VP-ellipsis. *Glossa: A Journal of General Linguistics*, 4(1). 60. 1–22. DOI: https://doi.org/10.5334/gjgl.738

Runner, Jeffrey T. & Head, Kellan D. L. 2014. What can visual world eye-tracking tell us about the Binding Theory? In Piñón, Christopher (ed.), *Empirical issues in syntax and semantics* 10, 269–286. http://www.cssp.cnrs.fr/eiss10/

Sag, Ivan. 1976. *Deletion and logical form*. Cambridge, MA: MIT dissertation.

Sag, Ivan & Hankamer, Jorge. 1984. Towards a theory of anaphoric processing. *Linguistics and Philosophy* 7. 325–345. DOI: https://doi.org/10.1007/BF00627709

Scuitti, Sandro. 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 Bardel, Camilla & Sánchez, Laura (eds.), *Third language acquisition: Age, proficiency and multilingualism*, 111–144. Berlin: Language Science Press.

Shapiro, Lewis P. & Hestvik, Arild. 1995. On-line comprehension of VP-ellipsis: Syntactic reconstruction and the semantic influence. *Journal of Psycholinguistic Research* 24. 517–532. DOI: https://doi.org/10.1007/BF02143165

Shapiro, Lewis P. & Hestvik, Arild & Lesan, Lesli & Garcia, A. Rachel. 2003. Charting the time-course of VP-ellipsis sentence comprehension: Evidence for an initial and independent structural analysis. *Journal of Memory and Language* 49. 1–19. DOI: https://doi.org/10.1016/S0749-596X(03)00026-3

Sprouse, Jon & Messick, Troy & Bobaljik, Jonathan David. 2022. Gender asymmetries in ellipsis: An experimental comparison of markedness and frequency accounts in English. *Journal of Linguistics* 58(2). 345–379. DOI: https://doi.org/10.1017/S0022226721000323

Storoshenko, Dennis Ryan. 2017. Mismatched gender in elided *self*-pronouns and *one*- anaphora. Poster presented at the Canadian Linguistic Association, May 27–29 2017.

Sudo, Yasutada & Spathas, Giorgos. 2020. Gender and interpretation in Greek: Comments on Merchant 2014. *Glossa* 5(1). 129. 1–42. DOI: https://doi.org/10.5334/gjgl.1173

Thomas, Margaret. 1989. The interpretation of English reflexive pronouns by non-native speakers. *Studies in Second Language Acquisition* 11. 281–303. DOI: https://doi.org/10.1017/S0272263100008147

Thomas, Margaret. 1991. Universal Grammar and the interpretation of reflexives in a second language. *Language* 67. 211–239. DOI: https://doi.org/10.1353/lan.1991.0065

Thomas, Margaret. 1993. *Knowledge of reflexives in a second language*. Amsterdam: John Benjamins Publishing Company. DOI: https://doi.org/10.1075/lald.6

Thoms, Gary. 2010. Tolerated lexical mismatches in ellipsis. Glasgow: Strathclyde University Ms.

Thoms, Gary. 2013. Lexical mismatches in ellipsis and the identity condition. In Keine, Stefan & Sloggett, Shayne (eds.), *Proceedings of the forty-second annual meeting of the North East Linguistics Society* (NELS 42), 559–572. Amherst, MA: GLSA Publications.

van Craenenbroeck, Jeroen. 2010. *The syntax of ellipsis: Evidence from Dutch dialects*. New York: Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780195375640.001.0001

Webber, Bonnie. 1979. *A Formal Approach to Discourse Anaphora*. Outstanding Dissertations in Linguistics. New York: Garland.

White, Joanna & Muñoz Lahoz, Carmen & Collins, Laura. 2007. The his/her challenge: Making progress in a 'regular' L2 programme. *Language Awareness* 16(4). 278–299. DOI: https://doi.org/10.2167/la419.0

Williams, Edwin. 1977. Discourse and logical form. *Linguistic Inquiry* 8(1). 101–39.

Xiao, Feng. 2015. Proficiency effect on L2 pragmatic competence. *Studies in Second Language Learning and Teaching* 5(4). 557–581. DOI: https://doi.org/10.14746/sllt.2015.5.4.3

Ying, Hongguang G. 2005. Second language learners' reading of reflexive anaphora in VP-ellipsis: A relevance theory perspective. *Language Sciences* 27. 551–570. DOI: https://doi.org/10.1016/j.langsci.2005.02.003

Yip, Virginia & Tang, Gladys. 1998. Acquisition of English reflexive binding by Cantonese learners: Testing the positive transfer hypothesis. In Beck, Maria-Luise (ed.), *Morphology and its interfaces in second language* knowledge, 165–193. Amsterdam: John Benjamins. DOI: https://doi.org/10.1075/lald.19.08yip

Yoshida, Mami. 2008. Think-aloud protocols and type of reading task: The issue of reactivity in L2 reading research. In Bowles, Melissa & Foote, Rebecca & Perpiñán, Silvia & Bhatt, Rakesh (eds.), *Selected proceedings of the 2007 second language research forum*, 199–209. Somerville, MA: Cascadilla Press.