Gender and ellipsis revisited

Rodrigo Ranero*

Abstract. Across several languages that encode grammatical gender, an intriguing pattern emerges under ellipsis. Whereas certain noun pairs disallow gender mismatches altogether between the antecedent and the ellipsis site, a second set of noun pairs allows them freely, and a third set allows them only when the grammatically masculine noun is in the antecedent but not vice versa (Bobaljik & Zocca 2011). We illustrate this pattern through the lens of Spanish and argue that the empirical generalizations can be captured via a universal identity condition regulating ellipsis that is split into two statements, where each statement refers to different syntactic primitives. On the one hand, the identity condition requires featural non-distinctness, which is a weaker requirement than strict featural identity. On the other hand, the identity condition requires that √ROOTs, unlike features, be strictly identical. Coupled with the independently needed mechanism of repair-by-ellipsis, we argue that the proposed identity condition can provide insight into capturing the microvariation that is attested across languages, within languages, and between individual grammars: whereas the identity condition does not vary, the featural representation of nominals varies idiosyncratically. We elaborate on the relevance of repair-by-ellipsis for this empirical domain and the identity condition, arguing that certain lexical gaps cannot be repaired (Mendes & Nevins 2022).

Keywords. ellipsis; gender; identity condition; repair; Spanish; microvariation

1. Introduction. Let us assume that ellipsis is regulated by a *universal* identity condition. In other words, whenever a silent expression is derived via the suppression of structure, that structure must satisfy an identity requirement with a discourse antecedent. With this assumption in mind, we can describe and analyze an array of grammars with the goal of proposing a single identity condition that can account for the ill-formed status of elliptical configurations for which identity violations underpin the deviance. For example, we should arrive at single condition that can account for this empirical observation: whereas all voice mismatches in English are ill-formed under sluicing (Merchant 2013), only a subset of voice mismatches are ill-formed in the Mayan languages Kaqchikel and Chuj (Ranero 2023; Ranero & Royer to appear).

Many different flavors have been proposed for the identity condition, some arguing that it is wholly semantic in nature, others arguing that it is wholly syntactic, and still others arguing that a mélange of both s-sides is required (see Merchant 2019; Ranero 2021: Ch. 1 for discussion). Here, we argue for the following formulation of the identity condition:

(1) Identity Condition on Ellipsis
   a. The antecedent and material properly contained in the ellipsis site must be featurally non-distinct.
   b. There must be a strict one-to-one match between all √ROOTs properly contained in the ellipsis site and √ROOTs in the antecedent.

* My heartfelt thanks to Masha Polinsky for her unwavering support! Thank you to two anonymous reviewers for detailed suggestions that led to improvements. This paper synthesizes aspects of Chapter 4 of Ranero (2021). Any mistakes are my own. Author: Rodrigo Ranero, University of California, Los Angeles (ranero@ucla.edu).

The empirical landscape discussed in what follows that will support (1) is the complex pattern under ellipsis of the nominal feature traditionally termed "gender". What we will observe is that there exist three different mismatch behaviors for noun pairs under ellipsis (see Bobaljik & Zocca 2011), where each member of the pair bears a different featural specification for the [GENDER] feature. We argue that the proposed condition in (1) and the independently needed mechanism of repair-by-ellipsis of morphophonological gaps can capture the empirical generalizations. Additionally, we argue that our approach can shed light on how to best capture the range of microvariation that has been reported, an issue that has seldom been tackled explicitly in previous work discussing our empirical puzzle. In a nutshell, the identity condition is invariant, but the lexical representation of nominals varies idiosyncratically cross- and intra-linguistically. These representational differences interact with the identity condition in a way that results in the three patterns that are observed.

2. The data. This section lays out how features in the nominal domain behave under ellipsis. Crucially, we observe that [GENDER] mismatches do not pattern uniformly. Instead, whether a mismatch is well-formed or ill-formed is contingent on noun "class" membership. Across languages, there seem to be three distinct noun classes, where for each class, [GENDER] mismatches are well-formed or ill-formed in a way that is distinct from the way mismatches behave in the other classes.

To contextualize the scope of the puzzle posed by [GENDER] in this domain of the grammar, consider first a different feature as a point of comparison. Observe how the English NP-ellipsis (NPE) examples in (2) show that [NUMBER] mismatch is well-formed. Notice from a methodological standpoint that the mismatch is forced within the ellipsis site by manipulating the remnant outside of the silence. In this case, singular or plural agreement on the remnant verb is manipulated to force the mismatch (the ellipsis site is notated through < > brackets moving forward):

(2) [NUMBER] mismatches are invariably well-formed
  a. Trixie's wig is on the stage and Katya's wigs are under the stage.
  b. Trixie's **wig** is on the stage and Katya's <**wigs**> are under the stage. ✓SG-PL
  c. Trixie's wigs are on the stage and Katya's wig is under the stage.
  d. Trixie's **wigs** are on the stage and Katya's <**wig**> is under the stage. ✓PL-SG

We return to the analysis of [NUMBER] mismatches and the relevance of their well-formed status in sections 3.2-3.3. For now, keep in mind that this configuration seems to be invariably well-formed (Saab 2019).[1] As alluded to before, though, other nominal features cannot mismatch freely. Let us zoom in on our focus here – grammatical gender – which we define as follows:

(3) Grammatical gender (Corbett 1991; see Kramer 2015: 65)
  Gender is the sorting of nouns into two or more classes as reflected in agreement morphology on determiners, verbs, and other syntactic categories.

We assume that grammatical gender distinctions are encoded on nominals via [GENDER] features in the little-n^0 head in the nominal spine, a formalization that we flesh out in section 3.1.

[1] Picallo (2017) argues that [NUMBER] mismatches in Spanish involving *pluralia tantum* nouns are ill-formed. However, a close examination of the crucial examples reveals that they are deviant even without ellipsis. Hence, the author's data do not bear on the identity condition whatsoever and we can maintain the generalization that [NUMBER] mismatch is invariably well-formed. See Ranero (2021: 329–334) for discussion.

Turning now to the behavior of [GENDER] under ellipsis, there exists a three-way contrast in behavior regarding well-formed and ill-formed mismatches in (at least) Brazilian Portuguese (Nunes & Zocca 2009; Bobaljik & Zocca 2011), Russian (Polinsky 2020), Greek (Merchant 2014; Alexiadou 2017; Spathas & Sudo 2020; Sudo & Spathas 2016, 2020), and Spanish (Saab 2010a; Donatelli 2019, among others); see section 6 for the issue of microvariation.[2] We illustrate this three-way pattern via Guatemalan Spanish; the empirical generalizations are summarized below:[3]

(4) [GENDER] mismatches are not uniform: three-way contrast
- a. Class I nouns: [GENDER] mismatch symmetrically ill-formed (6)–(7)
- b. Class II nouns: [GENDER] mismatch symmetrically well-formed (9)–(10)
- c. Class III nouns: [GENDER] mismatch asymmetrically well-formed (12)–(13)
  - (i) Grammatically masculine antecedent ✓
  - (ii) Grammatically feminine antecedent *

Note that the term "class" that we use throughout is meant to allow us to discuss each mismatch pattern under ellipsis in comparison to the others – whether each class has an internal logic independent of ellipsis is an issue deserving of more investigation (see footnote 15).[4]

2.1. CLASS I NOUNS. Let us begin with a sample of nouns belonging to Class I. Notice that these nouns denote kinship terms and animals, where the same root is shared by each pair member. Note as well that so-called "theme" vowels distinguish each pair member morphologically: *-o* for masculine and *-a* for feminine. Masculine nouns are listed first in each pair:

(5) Class I nouns

a. Kinship terms		b. Animals	
tío/tía	'uncle/aunt'	gato/gata	'cat'
primo/prima	'cousin'	cerdo/cerda	'pig'
suegro/suegra	'father-in-law/mother-in-law'	conejo/coneja	'rabbit'

As expected from the definition of grammatical gender in (3), masculine nouns are compatible with masculine determiners and masculine concord on adjectives; conversely, feminine nouns are compatible with feminine determiners and feminine concord on adjectives. We can use this fact about the reflexes of [GENDER] in the DP to assess mismatch possibilities under NPE. Baseline elliptical configurations with a [GENDER] match (6b,f) and target sentences with a [GENDER] mismatch (6d,h) are shown below. Observe how mismatches are forced through the remnant determiner and that mismatches are symmetrically ill-formed:

(6) Class I and NPE: symmetrically *

a. El gato de Max es dócil, pero el gato de Marta es feroz.
   the.M cat.M of Max is docile but the.M cat.M of Marta is ferocious
   'Max's (male) cat is docile, but Marta's (male) cat is ferocious.'

[2] See Barrie (2016) (Cayuga) and Murphy et al. (2018) (Bosnian-Croatian-Serbian) for related discussion. Bobaljik & Zocca (2011) report some German judgments as well, and a few noun pairs are given for German and Romanian that would fall under our Class I and III classifications.

[3] The data here represent the judgments of the author and three other native speakers of this dialect.

[4] We follow Merchant's (2014) numbering for the classes and their patterning. Unfortunately, it is impossible to be consistent with all authors in this regard – e.g., Bobaljik & Zocca (2011) and Donatelli (2019) deem noun pairs that allow mismatches symmetrically to be Class I (our Class II), those that allow mismatches asymmetrically Class II (our Class III), and those that disallow mismatches altogether Class III (our Class I).

b. El gato de Max es dócil, pero el <gato> de Marta es feroz. ✓M-M
'Max's (male) cat is docile, but Marta's (male cat) is ferocious.'

c. El gato de Max es dócil, pero la gata de Marta es feroz.
the.M cat.M of Max is docile but the.F cat.F of Marta is ferocious
'Max's (male) cat is docile, but Marta's (female) cat is ferocious.'

d. *El **gato** de Max es dócil, pero la <**gata**> de Marta es feroz. *M-F
Intended: 'Max's (male) cat is docile, but Marta's (female cat) is ferocious.'

e. La gata de Max es dócil, pero la gata de Marta es feroz.
the.F cat.F of Max is docile but the.F cat.F of Marta is ferocious
'Max's (female) cat is docile, but Marta's (female) cat is ferocious.'

f. La gata de Max es dócil, pero la <gata> de Marta es feroz. ✓F-F
'Max's (female) cat is docile, but Marta's (female cat) is ferocious.'

g. La gata de Max es dócil, pero el gato de Marta es feroz.
the.F cat.F of Max is docile but the.M cat.M of Marta is ferocious
'Max's (female) cat is docile, but Marta's (male) cat is ferocious.'

h. *La **gata** de Max es dócil, pero el <**gato**> de Marta es feroz. *F-M
Intended: 'Max's (female) cat is docile, but Marta's (male cat) is ferocious.'

Note furthermore that the size of the ellipsis site does not alter the pattern. In the predicate ellipsis examples below, we force the mismatch through the use of proper names associated with individuals who use masculine pronouns (e.g., Sebas) or feminine pronouns (e.g., Gaby). Note that a mismatch is symmetrically ill-formed here as well (7d,h):

(7) Class I and predicate ellipsis: symmetrically *
Context: Sebas, Pablo, Gaby, and Laura all have siblings who recently had children.

a. Sebas ya es tío y Pablo también ya es tío.
Sebas now is uncle and Pablo too now is uncle
'Sebas is an uncle now and Pablo is an uncle now too.'

b. Sebas ya es tío y Pablo también <ya es tío>. ✓M-M
'Sebas is an uncle now and Pablo is too.'

c. Sebas ya es tío y Gaby también ya es tía.
Sebas now is uncle and Gaby too now is aunt
'Sebas is an uncle now and Gaby is an aunt now too'

d. *Sebas ya es **tío** y Gaby también <ya es **tía**>. *M-F
Intended: 'Sebas is an uncle now and Gaby is (an aunt now) too.'

e. Laura ya es tía y Gaby también ya es tía.
Laura now is aunt and Gaby too now is aunt
'Laura is an aunt now and Gaby is an aunt now too.'

f. Laura ya es tía y Gaby también <ya es tía>. ✓F-F
'Laura is an aunt now and Gaby is (an aunt now) too.'

g. Laura ya es tía y Pablo también ya es tío.
Laura now is aunt and Pablo too now is uncle
'Laura is an aunt now and Pablo is an uncle now too.'

h. *Laura ya es **tía** y Pablo también <ya es **tío**>. *F-M
Intended: 'Laura is an aunt now and Pablo is (an uncle now) too.'

To summarize, we observe that [GENDER] mismatches are symmetrically ill-formed with Class I nouns irrespective of the elliptical construction under consideration.

2.2. CLASS II NOUNS AND ELLIPSIS. Class II nouns denote professions and occupations:

(8) Class II Nouns

abogado/abogada	'lawyer'	escritor/escritora	'writer'
maestro/maestra	'teacher'	(el) testigo / (la) testigo	'(the) witness'
jardinero/jardinera	'gardener'	(el) artista / (la) artista	'(the) artist'

Just like in Class I, each member of a noun pair in Class II bears the same root (observe the shared phonological profile between members of each pair). Note as well that a subset of these nouns is also morphologically parallel to Class I nouns in that the masculine bears the theme vowel *-o*, while the feminine bears *-a*. This isn't always the case, though, since occasionally the masculine bears no theme vowel at all ('writer') or both masculine and feminine pair-mates end in the same vowel ('artist', 'witness'). Morphologically, then, there is no clear-cut criterion that distinguishes these nouns from those in Class I.

Turning to ellipsis, the behavior of Class II nouns is the flipside of Class I: [GENDER] mismatches are symmetrically *well*-formed. Consider the NPE examples shown below (moving forward, we only show the crucial examples, bearing in mind that the baseline data – [GENDER] match, with and without ellipsis – are well-formed):

(9) Class II and NPE: symmetrically ✓

a. El abogado de Ana es competente y la abogada de Eu también.
the.M lawyer.M of Ana is competent and the.F lawyer.F of Eu too.
'Ana's (male) lawyer is competent and Eu's (female) lawyer is too.'

b. El abogado de Ana es competente y la <abogada> de Eu también. ✓M-F
'Ana's (male) lawyer is competent and Eu's (female lawyer) is too.'

c. La abogada de Ana es competente y el abogado de Eu también.
the.F lawyer.F of Ana is competent and the.M lawyer.M of Eu too.
'Ana's (female) lawyer is competent and Eu's (male) lawyer is too.'

d. ?La abogada de Ana es competente y el <abogado> de Eu también. ✓F-M
'Ana's (female) lawyer is competent and Eu's (male lawyer) is too.'

Judgments for (9d) are slightly degraded in comparison to (9b). However, the contrast with the behavior of Class I is stark – for nouns in Class I, any mismatch is impossible. The same difference in behavior between Class I and II is observable under predicate ellipsis:

(10) Class II and predicate ellipsis: symmetrically ✓

a. Matías no es abogado, pero Rosa sí es abogada.
Matías not is lawyer.M but Rosa yes is lawyer.F
'Matías is not a lawyer, but Rosa is a lawyer.'

b. Matías no es abogado, pero Rosa sí <es abogada>. ✓M-F
'Matías is not a lawyer, but Rosa is (a lawyer).'

c. Rosa no es abogada, pero Matías sí es abogado.
Rosa not is lawyer.F but Matías yes is lawyer.M
'Matías is not a lawyer, but Rosa is a lawyer.'

d. Rosa no es abogada, pero Matías sí <es abogado>. ✓F-M
'Rosa is not a lawyer, but Matías is (a lawyer).'

To summarize the empirical landscape thus far, Class I stands in clear contrast to Class II. Whereas the former disallows [GENDER] mismatches symmetrically, the latter allows them.

2.3. CLASS III NOUNS AND ELLIPSIS. Finally, we turn to Class III. Once again, note that a root appears to be shared within each pair. However, one morphological trait of Class III that is unique and will become relevant in section 4 is that the feminine noun in each pair bears a morpheme that is not borne by its masculine pair-mate (e.g., the apparent suffix *-sa* in *alcaldesa*):

(11) Class III Nouns
actor/actriz 'actor/actress'
alcalde/alcaldesa 'mayor'
poeta/poetisa 'poet'
héroe/heroína 'hero/heroine'

Turning to ellipsis, Class III exhibits a behavior that is distinct from both Class I and Class II. Zooming in on NPE first, observe that mismatches are allowed in an asymmetrical fashion: a masculine noun is allowed in the antecedent (12b), but a feminine one is not (12d). Note that we are assuming in this description that the elided noun in (12b) and (13d) is <actriz>, but we propose a different analysis in section 4 that will derive the asymmetrical behavior observed:

(12) Class III and NPE: ✓M-F /*F-M
a. El actor de Hollywood ganó un Óscar y la actriz de Bollywood también.
the.M actor of Hollywood won an Oscar and the.F actress of Bollywood too
'The Hollywood actor won an Oscar and the Bollywood actress did too.'
b. ?El actor de Hollywood ganó un Óscar y la <actriz> de Bollywood también. ✓M-F
'The Hollywood actor won an Oscar and the Bollywood (actress) did too.'
c. La actriz de Bollywood ganó un Óscar y el actor de Hollywood también.
the.F actress of Bollywood won an Oscar and the.M actor of Hollywood too
'The Bollywood actress won an Oscar and the Hollywood actor did too.' *F-M
d. *La actriz de Bollywood ganó un Óscar y el <actor> de Hollywood también.
Intended: 'The Bollywood actress won an Oscar and the Hollywood (actor) did too.'

We draw the same conclusion with predicate ellipsis:

(13) Class III and NPE: ✓M-F /*F-M
a. Jaime no es actor, pero Elizabeth sí es actriz.
Jaime not is actor but Elizabeth yes is actress
'Jaime is not an actor, but Elizabeth is an actress.'
b. ?Jaime no es actor, pero Elizabeth sí <es actriz>. ✓M-F
'Jaime is not an actor, but Elizabeth is (an actress).'
c. Laura no es actriz, pero Marlon sí es actor.
Laura not is actress but Marlon yes is actor
'Laura is not an actress, but Marlon is an actor.'
d. *Laura no es actriz, pero Marlon sí <es actor>. *F-M
Intended: 'Laura is not an actress, but Marlon is (an actor).'

To summarize, Class III exhibits an asymmetrical mismatch behavior that is distinct from the patterns for Class I and Class II: a mismatch is well-formed if the grammatically masculine noun is in the antecedent, but a mismatch is ill-formed if the feminine is in the antecedent.

2.4. EMPIRICAL SUMMARY AND KEY TAKEAWAY. The empirical picture that emerges for [GENDER] under ellipsis is complex –three distinct mismatch behaviors are observed under ellipsis in the

Spanish data assessed (4), and this pattern is replicated cross-linguistically (with microvariation that must also be accounted for; see section 6).

We propose that several ingredients are needed to account for the empirical generalizations we have established. First, a fine-grained conception of the featural composition of nominals in languages that encode [GENDER]. Second, the mechanism of repair-by-ellipsis. With these two ingredients in mind, the identity condition laid out at the outset in (1) becomes the third ingredient and can be implemented explicitly. We now turn to each of these.

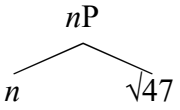
3. Analytical ingredients. This section lays out all the necessary components for our derivation of the intricate data described before. We begin in section 3.1 with the structure of nominals and licensing conditions for [GENDER]. Moving on to ellipsis, our assumptions regarding its derivation are laid out in section 3.2. The mechanism of repair-by-ellipsis and its limitations are laid out in section 3.3. Finally, the identity condition we proposed is discussed in finer detail in section 3.4.

3.1. NOMINAL DECOMPOSITION AND LICENSING CONDITIONS FOR [GENDER]. The locus of [GENDER] within the nominal spine has been a fertile area of research recently (see Mathieu et al. 2019; Kramer 2020, and references therein). Here, we adopt one specific proposal regarding the locus of this feature and assess how this approach can give us a handle on the complex empirical picture we established regarding mismatches under ellipsis.[5]

We assume a decompositional view of nominals withing the Distributed Morphology (DM) framework (Halle & Marantz 1993; Embick & Noyer 2007, among others). Concretely, we adopt the proposal that nominals are composed of several layers of structure where the innermost element is an acategorial √ROOT that is individuated syntactically via an index that instructs the interpretive (Encyclopedia) and externalization (Vocabulary Insertion) components post-syntactically (see Harley 2014; Kramer 2015; for proposals that take √ROOTs to be phonologically individuated, see Borer 2005). This approach posits, then, that √ROOTs are distinct syntactic primitives from features. We also adopt the view that √ROOTs combine with nominalizing little-n^0 heads within nominals.

Consider as an illustration the derivation of the noun 'mimosa' and the post-syntactic instructions for dealing with this structure: Vocabulary Insertion is concerned with externalization, whereas the Encyclopedia is concerned with interpretation. Our formalization follows Kramer (2015):[6]

(14) √ROOT categorization



Vocabulary Insertion: √47 ↔ [mɪ.ˈmoʊ.sa]
Encyclopedia: [$_{n\text{P}}$ [*n* [√47]] is interpreted as a flower, used in perfumery…

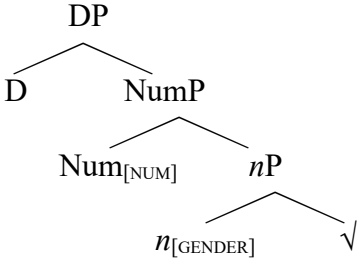
[5] An issue we do not discuss in detail is how [GENDER] features come to have reflexes on nominal elements such as determiners and adjectives (i.e., there exists gender concord in the DP). In short, we assume that concord is the reflex of a syntactic operation that is distinct from Agree and does not involve feature-sharing (see Norris 2014 for justification). One possible implementation is that concord involves feature percolation (Norris 2014); another is that it involves φ-feature copying (Polinsky 2016) – see Ranero (2021: 262–264).

[6] Instructions to the Encyclopedia could be notated as insertion as well (see, e.g., Mendes & Nevins' 2022 formalization in (25)). The choice is immaterial for our purposes, so we follow Kramer for consistency.

(i) Encyclopedia: √47 ↔ "a flower, used in perfumery…" / [$_{n\text{P}}$ [n _]]

We assume that [GENDER] is located on the nominalizing n^0 head (Kramer 2015). Furthermore, we assume that [NUMBER] is one layer above on Num^0 (Ritter 1991; see Preminger 2020) and the topmost later is D^0 (Abney 1987, among others):

(15) Full nominal structure



A key assumption that will play a role in our analysis is that in languages that encode grammatical gender, n^0 comes in different flavors, where each n^0carries a distinct [GENDER] specification. For Spanish, we assume the following set of n^0 heads (Kramer 2015):

(16) [GENDER] in Spanish
- a. The locus of [GENDER] is n^0.
- b. Default grammatical gender is masculine.
- c. There are 4 flavors of n^0:
  - (i) $n_{i[+F]}$ (interpretable feminine)
  - (ii) $n_{i[-F]}$ (interpretable masculine)
  - (iii) n (default; triggers masculine concord)
  - (iv) $n_{u[+F]}$ (uninterpretable feminine)

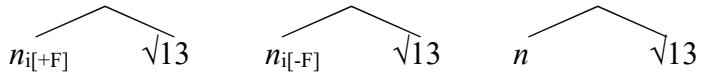
We assume that $_i$[+F] and $_i$[-F] contribute to interpretation: in the case of human denoting nouns, these features contribute an interpretation regarding pronoun use; in the case of animals, sex. The three-way opposition between (16c,i) feminine n^0 bearing an interpretable i[+F], (16c,ii) masculine n^0 bearing an interpretable i[-F], and (16c,iii) default n^0 allows us to capture three-way oppositions in Spanish, such as in the demonstrative system (e.g., *aquel* 'that' masculine, *aquella* 'that' feminine, *aquello* 'that'). Finally, positing (16c,iv) n^0 bearing an uninterpretable u[+F] allows us to distinguish pairs of inanimate nouns that trigger masculine or feminine concord but for which no contribution is made to the denotation by the [GENDER] feature: for example, *el suelo* 'the ground', categorized by (16c,iii); *la suela* 'the sole' categorized by (16c,iv). We briefly turn to inanimate nouns in section 5 and don't consider this flavor of n^0 until then.

Let us now turn in more detail to the interaction between n^0 and √ROOTs. We follow Kramer's (2015) proposal that √ROOTs are *licensed* by flavors of n^0 – i.e., there exist *licensing conditions* on nominals. These licensing conditions are a formalization of the restrictions on the compatibility of flavors of n^0 and specific √ROOTs. Take for instance a pair of Class I nouns like *tía/tío* 'aunt'/'uncle'. We propose that the √ROOT underlying these nominals can be licensed by feminine n^0 (17a) or masculine n^0 (17b), but cannot be licensed by default n^0 (17c) (notated *):[7]

[7] The source of the theme vowels *-o* and *-a* observed on Class I and II nouns cannot be developed in detail due to space and we abstract away from them for ease of exposition. Suffice to say that there exist tendencies, but no absolutes, regarding which ending appears on nouns of different grammatical genders (Harris 1991, among others). See Kramer (2015: 235) for an approach that is compatible with the assumptions adopted here.

(17) Licensing conditions on tía ‘aunt’ / tío ‘uncle’ (Class I)

a. *n*P b. *n*P c. * *n*P



(18) Post-syntactic instructions for (17a) and (17b)

a. Vocabulary Insertion: √13 ↔ [ˈti.a]
Encyclopedia: [$_{n\text{P}}$ [$n_{i[+F]}$ [√13]] is an individual, who is a parent’s sister, who…

b. Vocabulary Insertion: √13 ↔ [ˈti.o]
Encyclopedia: [$_{n\text{P}}$ [$n_{i[-F]}$ [√13]] is an individual, who is a parent’s brother, who…

We assume that licensing violations such as (17c) arise because of a problem in the Encyclopedia. In other words, there is no output for the interpretation of the relevant structure:

(19) Source of licensing violation for √13 (17c)
Encyclopedia: [$_{n\text{P}}$ [*n* [√13]] is interpreted as **???**

We further assume that in addition to licensing violations of the above type, an issue could arise in another post-syntactic module related to externalization (Vocabulary Insertion). In this case, there is no problem interpreting a combination of n^0 head and √ROOT (the source of a licensing violation). Rather, the structure cannot be pronounced – in other words, there exists a *morphophonological gap*. To illustrate abstractly, imagine a hypothetical √ROOT denoting an animal that can be licensed by any flavor of n^0, but for which an issue arises in the Vocabulary Insertion component when, e.g., combining with default n^0:

(20) Morphophonological gap for hypothetical √22

a. Vocabulary Insertion: √22 ↔ [ˈga.va.gai]
Encyclopedia: [$_{n\text{P}}$ [$n_{i[+F]}$ [√13]] is a female animal that runs around…

b. Vocabulary Insertion: √22 ↔ **???**
Encyclopedia: [$_{n\text{P}}$ [*n* [√13]] is an animal, that runs around…

Kramer (2015) discusses the kind of issue in (20b) as a violation of “arbitrary” licensing conditions, noting that this is the kind of violation that is usually discussed in the literature (see Harley & Noyer 1999; Siddiqi 2009). To avoid confusion, we use the term “morphophonological gap” for an issue like (20b), which we return to in more detail in section 3.3.

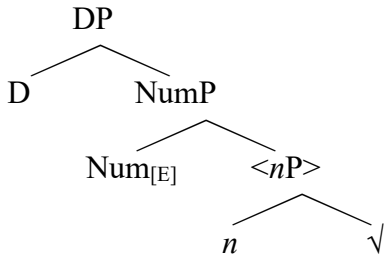
To summarize, we laid out the structure of the nominal spine, positing that its innermost element is an acategorial √ROOT, [GENDER] features reside on the nominalizing n^0 head, and [NUM] resides one layer above on Num^0. We also followed Kramer (2015) in proposing four distinct flavors of n^0 for Spanish. Finally, we laid out how n^0 interacts with acategorial √ROOTs – i.e., what licensing conditions on nominals are – discussing along the way two issues that could arise in the post-syntactic component of the grammar: (i) licensing violations that lead to uninterpretability in the Encyclopedia and (ii) morphophonological gaps that result from impossibility at Vocabulary Insertion. We now turn to the derivation of the elliptical constructions at the heart of this paper, followed by repair-by-ellipsis, and finally the identity condition.

3.2. THE DERIVATION OF ELLIPSIS. Let us be explicit regarding the derivation of the elliptical constructions that form the core of this paper. We adopt the viewpoint that ellipsis is deletion/non-insertion that is licensed in the syntax by a head bearing an [E] feature (Merchant 2001) that is compatible only with specific heads on a language-particular basis. For example, certain languages have auxiliary stranding VP-ellipsis, whereas others do not – this can be formalized by

positing that Voice can bear [E] in English (which possesses this ellipsis type), but Voice cannot bear [E] in other languages (Merchant 2013). The [E] feature instructs its complement for deletion/non-insertion (see Aelbrecht 2010; van Croenenbroeck & Merchant 2013 for elaboration), and the elided structure must satisfy an identity condition, which we turn to in section 3.4.

Let us pivot to the elliptical constructions that underlie the data here. We observed cases of predicate ellipsis with a remnant such as *también* 'too' or polarity particles such as *sí/no* 'yes/not' in (7), (10), and (13). We refer the reader to Brucart & McDonald (2012), who treat these examples as instances where the Σ^0 head bears [E] and TP is elided (see also Depiante 2004; Saab 2010b). Turning to NPE examples such as (6), (9), and (12), we follow Saab (2010a) in proposing that this configuration involves ellipsis of *n*P: Num^0 bears [E] and licenses ellipsis of its complement. For a discussion of diagnostics (e.g., sub-extraction asymmetries) that support the derivation of NPE as a surface anaphor/ellipsis, instead of a deep anaphor (using Hankamer & Sag's 1976 terminology), see Saab (2019) and Ranero (2021: 245–249).

(21) Derivation of NPE



The analysis above straightforwardly derives why [NUM] can mismatch under ellipsis – this feature is on Num^0, outside of the ellipsis site in NPE (Saab 2010a, 2019). It therefore does not enter into the calculation of compliance with the identity condition.

3.3. REPAIR-BY-ELLIPSIS. Starting with the seminal Ross (1969), it has been observed that certain ill-formed configurations appear to be well-formed under ellipsis.[8] In other words, ellipsis appears to "repair" a subset of grammatical violations (the term *salvation by deletion* is also used).

We follow Mendes (2020) and Mendes & Nevins (2022) and adopt the viewpoint that deviances with a morphophonological underpinning can be repaired by ellipsis, given that ellipsis is the literal absence of externalization. In other words, an issue that relates to the pronunciation of structure is (naturally) circumvented if said structure is suppressed. On the other hand, deviances that are underpinned by violations in the syntactic or semantic components cannot be circumvented by ellipsis, given that such deviances are unrelated to externalization proper.

Let us illustrate the repair capability of ellipsis via an example of a morphophonological gap in Guatemalan Spanish. Observe how the present tense paradigm for the verb *abolir* 'to abolish' exhibits several gaps – speakers cannot produce any form for some cells (Arregi & Nevins 2014; see Mendes & Nevins 2022 for Portuguese and Abels 2019b; Adamson 2019: 204):

[8] Ross discussed how sluicing (clausal ellipsis with a *wh*-remnant) appears to repair island violations. The existence of island repair has fueled much debate and we set it aside here; see Merchant (2001), Lasnik (2001), and Abels (2019a).

(22) Morphophonological gaps: Guatemalan Spanish

1SG	yo	-	Intended: 'I abolish.'
2SG	vos	abolís	'you abolish' (informal)
2SG	usted	-	Intended: 'you abolish' (formal)
3SG	élla/él	-	Intended: 's/he abolishes'
1PL	nosotros	abolimos	'we abolish'
2PL	ustedes	-	Intended: 'you abolish'
3PL	ellas/ellos	-	Intended: 'they abolish'

Conceivable forms for the gaps are judged unacceptable, e.g., **Yo abolo/abuelo*. Strikingly, however, the existence of a gap is tolerated under ellipsis (23b):

(23) Repair-by-ellipsis

a. *Vos abolís las leyes por interés, pero yo no abuelo/abolo las leyes por interés.
you abolish the laws bc. interest but I not abolish the laws bc. interest
Intended: 'You abolish the laws because of self-interest, but I don't abolish the laws because of self-interest.'

b. Vos abolís las leyes por interés, pero yo no < >.
'You abolish the laws because of self-interest, but I don't.'

We follow Mendes & Nevins (2022) in analyzing the repair effect as follows: the gaps in paradigms such as (22) are the result of a morphophonological issue. Namely, there is no Vocabulary Insertion output for certain combinations of φ-features, tense, and a subset of √ROOTs. Since ellipsis is the lack of externalization, though, we can naturally explain why said combinations are licit in examples like (23b). This repair capability of ellipsis will play a role in our account of the well-formed [GENDER] mismatches for Class III in section 4.3.

However, it has been noted that ellipsis cannot repair *all* grammatical deviances: for example, it appears that ellipsis cannot circumvent Superiority violations (Boeckx & Lasnik 2006), ECP violations (Nakao 2009), or HMC violations (Mendes 2020); see also Mendes & Kandybowicz (2023) on extraction from perfect clauses in Nupe. Of crucial interest to us are certain configurations reported by Mendes & Nevins (2022). Consider the idiomatic expression *high jinks* meaning 'mischief', which cannot occur in the singular (**high jink*) and cannot be altered to involve another adjective; e.g., **low jinks*. Mendes & Nevins report that attempting to elide a singular version of this expression is ill-formed. Note that this manipulation is available testing ground since number mismatches are well-formed under ellipsis (recall (2); Saab 2019):

(24) No repair-by-ellipsis
*I don't care for these high jinks, not even one < >.

To account for the contrast between the above ill-formed example and the well-formed status of examples like (23b), Mendes & Nevins propose that the issue with a singular version of *high jinks* does not lie in Vocabulary Insertion, but in the *Enyclopedia*. Following the authors' notation, observe how the Encyclopedia can only interpret the root underlying the expression (√JINK) in the single structural context of [+PLURAL] and the adjective *high* (here in Spec, DP):

(25) Encyclopedia entry for √JINK (notation from Mendes & Nevins 2022: 7)
√JINK ↔ mischief′ / [$_{DP}$ high [$_{\#P}$ [$_{n P}$ [_ *n*] [+PLURAL]]] *no elsewhere item*

Given that the example in (25) would lead to an issue in the Encyclopedia – i.e., there is simply no interpretation available for √JINK in the context of singular – then ellipsis cannot

circumvent the issue. The absence of repair under ellipsis for deviances and gaps unrelated to externalization is also discussed when analyzing the [GENDER] mismatch patterns that are our focus and when we expand the empirical picture in section 5.

To summarize this sub-section, we presented evidence that certain grammatical deviances can be repaired by ellipsis – namely, lexical gaps that have to do with morphophonological externalization – whereas others that are unrelated to externalization cannot be repaired.

3.4. THE IDENTITY CONDITION. We repeat below our proposed identity condition:

(26) Identity condition on ellipsis (repeated from (1))
   a. The antecedent and material properly contained in the ellipsis site must be featurally non-distinct.
   b. There must be a strict one-to-one match between all √ROOTs properly contained in the ellipsis site and √ROOTs in the antecedent.

The bipartite nature of (26) distinguishing the syntactic primitive of *features* on the one hand (26a) and √*ROOTs* on the other (26b) is directly inspired by Saab's (2008: 36) identity condition. Saab's proposal differs, however, in requiring strict identity of features (for other recent proposals requiring strict identity, see Merchant 2013; Rudin 2019).

Let us develop each of the statements composing (26). The idea that non-distinctness is at play in the derivation of ellipsis is first found in Chomsky's (1965) discussion of comparatives (see Lipták 2013). What is needed for our present purposes is establishing the configurations that are ruled-in by non-distinctness and the configurations that are ruled out. In a nutshell, featural non-distinctness is violated when there exists a *featural clash*. Conversely, a mismatch between a featurally specified node and an equivalent node lacking the relevant features is allowed. Furthermore, a configuration in which the antecedent or ellipsis site lack the relevant node altogether is also ruled in. This is all schematized below:

Antecedent	Ellipsis site	Status
$H_{[F: X]}$	$H_{[F: Y]}$	*
$H_{[F: Y]}$	$H_{[F: X]}$	*
$H_{Ø}$	$H_{[F: X]}$	✓
$H_{[F: X]}$	$H_{Ø}$	✓
Ø	$H_{[F: X]}$	✓
$H_{[F: X]}$	Ø	✓

Table 1: Mismatches and featural non-distinctness (26a)

There exist empirical domains independent of our [GENDER] concerns here that support the proposal that featural non-distinctness plays a role in regulating ellipsis. Without delving into detail, these domains involve asymmetries in tense and polarity mismatches between languages like English and Spanish (Stockwell & Wong 2020; Ranero 2021) and asymmetries in voice mismatch between Mayan languages (Ranero & Royer to appear) and others (e.g., Merchant 2013).

Turning to √ROOTs, the proposed identity condition enforces a strict one-to-one match between the antecedent and the ellipsis site; i.e., this is a stricter condition, schematized in Table 2:

Antecedent	Ellipsis site	Status
√1	√2	*
Ø	√1	*
√1	Ø	*
√1	√1	✓

Table 2: Mismatches and strict √ROOT identity (26b)

There also exists independent evidence that a statement enforcing strict identity on √ROOTs is warranted. For example, this proposal gives us a handle on lexical identity effects in ellipsis (Rooth 1992; see Chung's 2006 generalization and Ranero 2021: 352 for discussion). A striking illustration for our purposes is found in Merchant (2019). Consider the synonyms *wedding* and *nuptials*, the latter being a strictly plural noun (**nuptial*). Armed with the knowledge that number mismatches are generally available in ellipsis (recall (2) and (24)), we can attempt to mismatch these two lexical items. However, such a mismatch is ill-formed. Given that the [NUMBER] specification could not be to blame here, the clash of √ROOTs must underpin the unacceptability:[9]

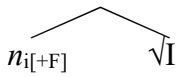
(27) √ROOT identity under ellipsis
  a. Beth's nuptials were in Bond Chapel, and Rachel's <nuptials> were in Rockefeller chapel.
  b. * Beth's **nuptials** were in Bond Chapel, and Rachel's <**wedding**> was in Rockefeller chapel.

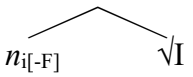
Having discussed our proposed identity condition in further detail, we are now ready to implement the analysis in the next section.

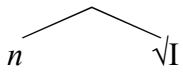
4. Implementing the analysis. Here, we lay out how the combination of analytical ingredients discussed previously can derive the three-way contrast in [GENDER] mismatch.

4.1. CLASS I MISMATCHES. Recall that Class I mismatches are ill-formed symmetrically. We propose the following LF licensing conditions for nouns in this class:

(28) Licensing conditions for Class I

a. *n*P


b. *n*P


c. **n*P


We can thus derive the ill-formed status of mismatches for Class I: they involve a clash of featurally specified n^0 heads. The mismatches thus violate featural non-distinctness (26a):

(29) [GENDER] mismatches in Class I

	Antecedent	Ellipsis site	Example	Status
a.	$n_{i[-F]}$	$n_{i[+F]}$	(6d), (7d)	*
b.	$n_{i[+F]}$	$n_{i[-F]}$	(6h), (7h)	*

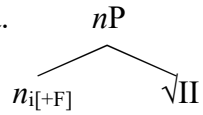
Recall, however, that we stated that masculine is the default gender in Spanish. A question arises, then: why isn't there a well-formed mismatch configuration for Class I where a feminine noun is in the antecedent and a default n^0 (controlling masculine concord) is in the ellipsis site?

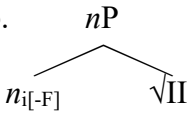
[9] An identical illustration for the strict √ROOT requirement can be made through Spanish by mismatching *heces* ('feces') – which is grammatically feminine and strictly plural – and *caca* (also grammatically feminine). We do not provide these data.

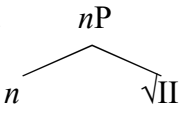
We rule out this overgeneration via our proposed licensing conditions in (28). Recall that we adopt the viewpoint that a licensing violation such as (28c) is the result of a problem in the Encyclopedia. Recall further that we adopt Mendes & Nevins' (2022) insight that issues unrelated to externalization – such as problems with *interpretation* – cannot be repaired by ellipsis. Thus, a configuration where there is a default n^0 in the ellipsis site *would* satisfy the identity condition but said configuration would incur in a licensing violation that cannot be repaired by ellipsis.

4.2. CLASS II MISMATCHES. In this case, we must derive the diametrically opposite pattern to Class I's—with Class II, mismatches are symmetrically well-formed. We thus propose the following licensing conditions for this set of nouns:

(30) Licensing conditions for Class II

a. 

b. 

c. 

Our proposal is that the freedom of licensing in this class renders the mismatches we constructed well-formed, since a default n^0 (30c) is involved (recall that this default nominalizer triggers masculine reflexes on determiners and modifiers in Spanish). The configurations below comply with featural non-distinctness – there is no featural clash:[10]

(31) [GENDER] mismatches in Class II

	Antecedent	Ellipsis site	Example	Status
a.	n	$n_{i[+F]}$	(9b), (10b)	✓
b.	$n_{i[+F]}$	n	(9d), (10d)	✓

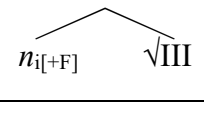
Consider at this juncture how we are able to account for grammars that generate these configurations because our condition does not enforce *strict* featural identity. Under approaches that *do* require strict identity, though, it is unclear how the pattern could be derived straightforwardly.

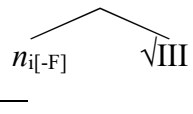
4.3. CLASS III MISMATCHES. The pattern with Class III has two characteristics that any account must derive. First, we must account for why a mismatch is well-formed when the grammatically masculine pair-mate is in the antecedent. Conversely, we must account for why a mismatch is ill-formed in the opposite configuration, where the feminine in the antecedent. Starting with the first side of the coin, what we propose is the following:

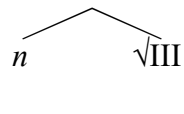
(32) Class III and [GENDER] mismatch: ✓ masculine-feminine
A mismatch where the masculine noun is in the antecedent and the feminine noun is in the ellipsis site is well-formed because of repair-by-ellipsis of a morphophonological gap.

Let us start to flesh out (32) by proposing the licensing conditions for Class III. Nouns in this class can be licensed by any of the following n^0 heads (see footnote 10):

(33) Licensing conditions for Class III (surface masculine)

a. 

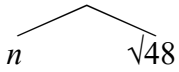
b. 

c. 

[10] There is some independent evidence that Class II and III nouns can be licensed under this default *n*: speakers accept a masculine noun as a neutral term (Harris 1991). However, judgments related to this appear to be in flux today, and speakers are often hesitant in judging them as well-formed. There is no such flexibility for Class I nouns though (e.g., **Laura es mi tío* 'Laura is my uncle', where the context is clear that Laura uses feminine pronouns).

What the proposal above entails is that nouns like *alcalde* '(male) mayor' can have the following structure instructing the post-syntactic modules:

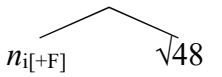
(34) Licensing conditions for Class III



Vocabulary Insertion: √48 ↔ [al.ˈkal.de]
Encyclopedia: [$_{n\text{P}}$ [*n* [√48]] is interpreted as a government official…

What about the post-syntactic instructions for a Class III √ROOT that is licensed by feminine n^0 (33b)? Our proposal is that this structural configuration is indeed well-formed on the s-sides, but there is a morphophonological gap – there are no instructions at Vocabulary Insertion (we return to the derivation of pronounceable forms such as *alcaldesa* and *actriz* below):

(35) Licensing conditions for Class III (lexical gap)



Vocabulary Insertion: √48 ↔ **???**
Encyclopedia: [$_{n\text{P}}$ [$n_{\text{i[+F]}}$ [√48]] is interpreted as a government official, who uses female pronouns…

Now we can zoom in on the well-formed mismatches in Class III. What is elided in these masculine-feminine mismatches is a configuration like in (36) – the antecedent noun is licensed by default n^0, while the ellipsis site includes a noun licensed by feminine n^0 (35). Thus, there is no featural clash and the identity condition is satisfied. Furthermore, the morphophonological gap in the ellipsis site is repaired by ellipsis. As a result, the example is well-formed:[11]

(36) [GENDER] mismatches in Class III: well-formed masculine-feminine

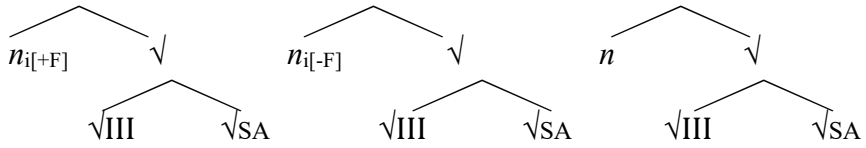
Antecedent	Ellipsis site	Example	Status
n	$n_{\text{i[+F]}}$ (VI gap)	(12b), (13b)	✓

We have now derived one side of the Class III puzzle. At this juncture, then, we need to turn to the other side of the puzzle and rule out examples where the (pronounceable) feminine noun in Class III is in the antecedent. We propose that the morphemes borne by feminine nouns in this class (e.g., *-sa*, *-iz, -ína*) are √ROOTs that cannot surface independently, but must combine with another √ROOT (Creemers et al. 2018; see Lowenstamm 2015).[12] In other words, these morphemes are akin to *cran*-morphs that cannot be freestanding. With this proposal in mind, observe our proposal for the licensing conditions for a feminine Class III noun like *alcaldesa* '(female) mayor': these nouns are only licensed by feminine n^0 (37a). For ease of illustration, we use √SA here instead of a √ROOT with a numerical index.

[11] Circling back to the data points, note that, e.g., <a*ctriz*> is *not* what is in the ellipsis site – instead, there is a gap.
[12] Our proposal is similar in spirit to Alexiadou's (2017) discussion of Greek, where she proposes that "derivational" suffixes akin to the morphemes discussed here are to blame for the ill-formed mismatches. Polinsky (2020) also notes the presence of these kind of "derivational" affixes for Class III in Russian (see also Bobaljik & Zocca 2011). Donatelli (2019: 194) sets aside this line of analysis for Spanish, appealing to a high degree of morphological dissimilarity for pairs like *actor/actriz*. We disagree that this criterion should lead us to reject our approach.

(37) Licensing conditions for Class III (surface feminine)

a. *n*P b. * *n*P c. * *n*P



As a result of the above proposal, we can derive the ill-formed status of a mismatch where a feminine noun in Class III is in the antecedent. The requirement for featural non-distinctness is violated in (38a), and both (38a-b) incur in licensing violations that cannot be repaired by ellipsis:

(38) [GENDER] mismatches in Class III: ill-formed feminine-masculine

	Antecedent	Ellipsis site	Example	Status
a.	[$n_{i[+F]}$ … [√III √IZ]]	[$n_{i[-F]}$ … [√III √IZ]]	(12d), (13d)	*
b.	[$n_{i[+F]}$ … [√III √IZ]]	[*n* … [√III √IZ]]	(12d), (13d)	*

Another configuration that could underlie the relevant (ill-formed) examples needs to be ruled out too: a feminine-masculine mismatch where the ellipsis site contains a form like *alcalde* 'mayor' or *actor* in which a single √ROOT is licensed by default n^0 (recall (34)). This configuration is correctly ruled out since it violates the requirement for strict √ROOT identity in our identity condition (26b): there is a √ROOT present in the antecedent that has no match in the ellipsis site:

(39) [GENDER] mismatches in Class III: ill-formed feminine-masculine

Antecedent	Ellipsis site	Example	Status
[$n_{i[+F]}$ … [√III <u>√IZ</u>]]	[*n* √III]	(12d), (13d)	*

To summarize our account of the asymmetrical mismatch behavior in Class III, we proposed that the well-formed mismatches are the result of repair-by-ellipsis of a configuration that satisfies the identity condition. Conversely, there is no configuration where the feminine is in the antecedent that satisfies identity and/or complies with our proposed licensing conditions.

4.4. SUMMARY. This section implemented the analytical ingredients laid out in section 3 and provided a stepwise account of how our three-way contrast in mismatch behavior can be derived. While a detailed comparison to competing approaches cannot be given due to space (see Ranero 2021: 304–314), we laid out how an identity condition predicated on featural non-distinctness and strict √ROOT matching can derive the pattern, alongside our assumptions about the featural decomposition of nominals and the mechanism of repair-by-ellipsis.[13]

5. Extending the empirical picture. We now extend the empirical picture by discussing noun pairs where there exists a gap in one of the cells – i.e., pairs in which there exists a gap for a grammatically masculine or for a grammatically feminine pair-mate. We provide evidence that none of the gaps can be repaired by ellipsis: all mismatch configurations are ill-formed. We take this to further support the idea that some lexical gaps, but not all, can be repaired (Mendes & Nevins 2022). We then turn to the behavior of nouns that denote inanimate entities.

[13] A reviewer asks how our proposal handles gender systems like Greek's, where the empirical picture concerning ellipsis is complicated by the availability of neuter gender, and thus a more complex inventory of nominalizing heads would be needed. For example, there is an interesting observation in Spathas & Sudo (2020: 39) that nouns denoting animals do not allow for gender mismatches under ellipsis even when one of the mismatching elements is neuter. We leave a derivation of these Greek data for the future due to space limitations.

Our first set of data is shown below. Note that we use NPE, but predicate ellipsis is also ill-formed. Note as well that baseline examples with ellipsis and a [GENDER] match are well-formed, but omitted due to space reasons:

(40) Extending the empirical picture: lexical gaps that cannot be repaired

a. *La vaca de Aída no es feliz, pero el <> de Ana sí es feliz.
the.F cow of Aída not is happy but the.M of Ana yes is happy
Intended: 'Aída's cow isn't happy, but Ana's (masculine counterpart) is.'

b. *El toro de Aída no es feliz, pero la <> de Ana sí es feliz.
the.M bull of Aída not is happy but the.F of Ana yes is happy
Intended: 'Aída's bull isn't happy, but Ana's (feminine counterpart) is.'

c. *Llegaron la institutriz de Ana y el <> de Elsa.
arrived the.F governess of Ana and the.M of Elsa
Intended: 'Ana's governess and Elsa's (masculine counterpart) arrived.'

d. *Llegaron el gramático de UMD y la <> de UCSB.
arrived the.M grammarian of UMD and the.F of UCSB
Intended: 'The UMD grammarian and UCSB's (female counterpart) arrived.'

e. *Se escaparon la culebra de Elena y el <> de Olga.
SE escaped the.F snake of Elena and the.M of Olga
Intended: 'Elena's snake and Olga's (masculine counterpart) escaped.'

f. *Se escaparon el delfín de Elena y la <> de Olga.
SE escaped the.M dolphin of Elena and the.F of Olga
Intended: 'Elena's dolphin and Olga's (masculine counterpart) escaped.'

g. *La víctima del asalto y el <> del fraude declaran hoy.
the.F victim of robbery and the.M of fraud declare today
Intended: 'The victim of the robbery and the fraud's (masc. victim) declare today.'

Consider first grammatically feminine nouns for which there is no masculine counterpart sharing the same √ROOT. These include kinship terms such as *nuera* 'daughter-in-law' and *madre* 'mother', as well and animals such as *vaca* 'cow', *yegua* 'mare' and *iguana* '(female) iguana'. Examining the behavior of these nouns under ellipsis is interesting, because a [GENDER] mismatch could be well-formed, leading us to conclude that the lack of a masculine counterpart is the result of a morphophonological gap in a configuration where default n^0 licenses the relevant √ROOTs. Conversely, the mismatch could be ill-formed, leading us to conclude instead that licensing violations underlie the gap. The result in (40a) indicates that the gap cannot be repaired – i.e., the source of the gap is a licensing violation, an issue unrelated to externalization. Put simply, this set of nouns must be licensed solely by feminine $n_{i[+F]}$ – i.e., the configuration in (40a) cannot involve a noun licensed by a *default* n^0 in the ellipsis site, which would lead to (i) identity being satisfied and (ii) a gap related to externalization being repaired, contrary to fact.

The nouns just described above have semantic counterparts: masculine nouns for which there is no feminine pair-mate that shares *the same* √ROOT. We thus find the kinship terms *yerno* 'son-in-law' and *padre* 'father', and the animals *toro* 'bull', *caballo* 'horse', and *garrobo* '(male) iguana'. The prior literature, in fact, has occasionally coupled nouns like *vaca/toro* as a natural class (e.g., Saab 2010a terms these "suppletive" pairs). In a similar manner to example (40a), we can test the behavior of this set of masculine nouns under ellipsis and force a [GENDER] mismatch. The result is ill-formed, suggesting that these nouns are licensed only by masculine $n_{i[-F]}$. In other words, the gap in (40b) cannot be repaired by ellipsis: i.e., this configuration (40b) could

not involve a noun licensed by a *default* n^0 in the antecedent, leading to (i) identity being satisfied and (ii) a gap related to externalization for a feminine $n_{i[+F]}$ in the ellipsis site being repaired.

Another set of nouns that (to our knowledge) has not been discussed in the literature concerns feminine nouns for which there is no masculine counterpart at all (where the masculine feature would have an interpretation regarding pronoun usage). One example is *institutriz* 'governess'. A [GENDER] mismatch manipulating this noun is also ill-formed (40c), suggesting that this nominal can only be licensed by a single nominalizing head – here, feminine $n_{i[+F]}$. We conclude that the relevant lexical gap here cannot be repaired by ellipsis either and is hence unrelated to the externalization component.

A few grammatically masculine nouns also lack a feminine counterpart – at least for some speakers – where the feminine feature would have an interpretation regarding pronoun usage. For example, some speakers only accept *músico* '(male) musician' and *gramático* '(male) grammarian', judging the feminine counterparts ill-formed.[14] Once again, a [GENDER] mismatch can be constructed but the result is unacceptable (40d), suggesting that these nominals are only licensed by masculine $n_{i[-F]}$ and there is no repair possibility.

Two examples above involve animal denoting nouns that are grammatically feminine (*culebra* 'snake') or masculine (*delfín* 'dolphin') for which any specification regarding sex must be stated via the use of the modifiers *macho* 'male' and *hembra* 'female'; e.g., *culebra macho* 'male snake' and *culebra hembra* 'female snake'. Starting with the grammatically feminine nouns, observe that a [GENDER] mismatch is ill-formed (40e). We propose that these nouns are licensed solely by a feminine nominalizing head bearing an uninterpretable feature $n_{u[+F]}$ – i.e., a feature that triggers feminine concord on modifiers but does not contribute to the denotation anything related to sex (in the case of animals; recall the flavors of n^0 we proposed for Spanish in (16)). Once again, we propose that the mismatch is ill-formed because of a licensing violation for the nominal in the ellipsis site. Turning to the grammatically masculine nouns, a mismatch is also ill-formed (40f). Here, we propose that these nouns are licensed only by default n^0 and the mismatch results in a licensing violation in the ellipsis site that cannot be repaired either.

Finally, consider nouns like *víctima* 'victim', *persona* 'person' and *criatura* 'creature', 'baby', which are grammatically feminine, but can be used to denote an individual of any gender. A [GENDER] mismatch is also ill-formed with these nouns (40g). We interpret this result as follows: these nouns are licensed solely by a female nominalizing head bearing an uninterpretable feature $n_{u[+F]}$, which does not contribute anything related to pronoun usage to the denotation. The mismatch in (40g) is ill-formed because of a licensing violation, and it cannot be repaired.

Before concluding this section, let us assess a final set of data involving nouns that denote inanimate entities; baseline examples without ellipsis are well-formed, but omitted ((41a) is adapted from Saab 2008):

[14] We are not proposing that there is no way to express the thought of a musician who uses feminine pronouns – rather, *música* is overwhelmingly interpreted as the abstract 'music', and an alternative strategy is thus employed to express the relevant thought, such as *Gaby hace música* 'Gaby does music'. Alas, only the author accepts *gramático* – other speakers who were consulted do not use this word productively at all.

(41) Extending the empirical picture: inanimates

a. *El orden natural de las cosas no puede ser alterado por una
the.M order natural of the things not can be altered.M by a.F
<orden> arbitraria de Dios.
command arbitrary.F of God
Intended: 'The natural order of things cannot be altered by God's arbitrary (command).'

b. *El cerezo es viejo, pero las <cerezas> de sus ramas son deliciosas.
the.M cherry.tree is old.M but the.F.PL cherries of its branches are delicious
Intended: 'The cherry tree is old, but the (cherries) of its branches are delicious.'

All [GENDER] mismatches with these kinds of nouns are ill-formed. We follow Saab's (2008) proposal to rule out these examples: the form-relatedness in these pairs is the result of partial or total homophony from a synchronic perspective – these pairs do not share the same √ROOT. In other words, identity here is violated because of a violation of strict √ROOT identity.

Let us summarize the takeaways from this section. First, the examples assessed in (40) suggested that there are configurations that satisfy the identity condition that nevertheless incur in licensing violations between n^0 and the √ROOT. These create an issue for the Encyclopedia that cannot be repaired by ellipsis. Second, any [GENDER] mismatch involving the manipulation of nouns denoting inanimate entities is ill-formed (41). We accounted for this by appealing to a violation of the statement requiring strict √ROOT matching in our proposed identity condition.

6. Conclusion: on microvariation and an open question. Let us end by highlighting how our analytical approach to [GENDER] mismatches in ellipsis is supported by its potential to account for the existence of microvariation along three different dimensions (42). Consider a remark in (43) by Merchant (2014) that serves to illustrate (42a-b):

(42) Dimensions of microvariation in [GENDER] mismatches

a. Cross-linguistic classification of noun pairs into Class X or Y
b. Intra-linguistic classification of noun pairs into Class X or Y
c. Difference in acceptability depending on elliptical construction (NPE vs. others)

(43) Greek

Lastly, some speakers vary in which class they assign a given pair to; the examples consist of cases where speakers were uniform, but the lists contain items that are true of at least one speaker (while others may differ: for example, though *thios/thia* 'uncle/aunt' is listed here in the one-way alternating class in accordance with the judgments of my primary informant, at least one speaker assigned it to the non-alternating class). A fuller exploration of the variation in this domain is needed. (Merchant 2014: fn. 6)

What Merchant is highlighting above is the existence of *cross-linguistic* and *intra-linguistic* variation regarding the categorization of the noun pair equivalent to the kinship term 'uncle/aunt'. We place this pair into Class I in Spanish and a subset of Greek speakers also classifies it in this way, but another subset classifies the pair instead as Class III.

Looking at Spanish in closer detail reveals more examples of intra-linguistic variation (42b). Consider how Donatelli (2019: 217) provides an example of a Class II pair in our terms: *médico/médica* 'physician'. The masculine-feminine mismatch is judged fully acceptable, but the author provides a judgment of ? for the feminine-masculine configuration, commenting that this example "is acceptable for the majority of speakers, though some slightly disprefer it to [the

masculine-feminine mismatch]". Donatelli goes on to conclude that some speakers treat this pair, then, like our Class III pair *actor/actriz*, which exhibits the asymmetrical mismatch behavior. Related conclusions could be drawn from remarks in Saab (2004: 51).

Furthermore, consider how Bobaljik & Zocca (2011) report in list form that Spanish noun pairs denoting animals like *gato/gata* 'cat' behave like our Class III, whereas we provided judgments that led us to place them in Class I (Saab 2008: 506 provides equivalent judgments). Bobaljik & Zocca (2011) also provide a Russian pair *moskovič/moscovička* 'Muscovite', reporting that it behaves like our Class III noun. Other Russian speakers, however, treat this pair as our Class II, allowing mismatches symmetrically (Maria Polinsky p.c.).

Overall, then, the picture that we find regarding microvariation (42a-b) should lead us to reflect on what kind of analysis is necessary to take this microvariation seriously and account for the totality of reported grammars. We contend that the one ingredient that should remain invariant is the identity condition – i.e., there should be no parameterization regarding this grammatical universal. Armed with this assumption, it becomes clear that the proposal laid out in this paper provides us with enough flexibility to make sense of microvariation (42a-b). In a nutshell, speakers across and within languages might have subtly different representations for the featural decomposition of equivalent nominal pairs. To be precise about an example discussed above, certain speakers of Russian might treat the *-ka* suffix in the grammatically feminine version of 'Muscovite' as a √ROOT, in a manner parallel to our account of why Class III mismatches in Spanish exhibit an asymmetrical behavior. If this is on the right track, then we can explain why Bobaljik & Zocca (2011) report that this noun is part of a Class III pair; conversely, other speakers do not represent *-ka* as a √ROOT, but as a morpheme akin to Spanish theme vowels that does not trigger an identity violation. Put simply, what we are proposing is that nominal representations can vary idiosyncratically, but the identity condition is invariant.

The final dimension that (might) exhibit microvariation concerns the type of elliptical construction under consideration (for fuller remarks, see Ranero 2021: 322). Unfortunately, the existence of variation along this dimension is more difficult to establish with certainty, given that previous authors rarely provide a full set of examples for all noun classes while controlling for elliptical construction. For example, one can establish that a tripartite behavior exists in Brazilian Portuguese, but Bobaljik & Zocca (2011) only discuss predicate ellipsis and relegate NPE to a footnote, while Nunes & Zocca (2009) discuss NPE, but the data is missing examples for our Class III that would establish the three-way contrast (these data are given in Ranero 2021: 227).

In contrast, one would be led to think that ellipsis type does matter in Greek: Merchant (2014) reports that this language showcases the three-way contrast exclusively in predicate ellipsis, whereas all [GENDER] mismatches are ill-formed in NPE. However, this empirical generalization was challenged by Sudo & Spathas (2016) and Alexiadou (2017), who argue that Merchant's NPE examples were ill-formed independently of the mismatches.

The literature on Spanish is difficult to assess with regards to (42c). For example, consider how Donatelli's (2019) study provides only two NPE examples (both ill-formed): the first involving a noun pair in our Class I and the second involving nouns that we would analyze as exhibiting an Encyclopedic gap. To our knowledge, the clearest exemplar of a grammar that might distinguish between ellipsis types is Saab (2008), who provides a well-formed judgment for a [GENDER] mismatch involving *maestro/maestra* 'teacher', but a ?? judgment for an NPE example involving *medico/médica* 'physician', pairs that we would place in Class II.

Regardless of the difficulty in establishing the existence of such a grammar through the literature, it is worth exploring if we *could* account for such a system. First, let us maintain that the

identity condition should be blind to the elliptical "construction" under consideration. Instead of appealing to any variation in the condition, then, we could take inspiration from the analytical ingredients used in this paper and from Merchant's (2014) interpretation of Greek data. Suppose, then, that some speakers are strict in their featural representations for [GENDER] and thus arrive at grammars that only have a Class I. Thus, any mismatch violates the identity condition. However, we can follow Merchant (2014) and wonder whether these speakers represent certain silent expressions as *deep anaphors*, instead of ellipsis (Hankamer & Sag 1976). Thus, any flexibility that they exhibit regarding [GENDER] alternations for those deep anaphors should not be analyzed via the narrow grammar, but rather grammar-externally – the identity condition would not be at play and pragmatic considerations would govern the pattern, considerations that would need to be explored. Given this approach – merely sketched and endowed with predictions – we could maintain our commitment to the invariant and universal identity condition in (1).[15]

References

Abels, Klaus. 2019a. Movement and islands. In Jeroen van Craenenbroeck & Tanja Temmerman (eds.), *The Oxford handbook of ellipsis,* 389–424. Oxford: Oxford University Press.

Abels, Klaus. 2019b. On "sluicing" with apparent massive pied-piping. *Natural Language and Linguistic Theory* 37. 1205– 1271. https://doi.org/10.1007/s11049-018-9432-1.

Abney, Steven P. 1987. *The English noun-phrase in its sentential aspect*. Cambridge, MA: MIT dissertation.

Adamson, Luke. 2019. *Derivational trapping and the morphosyntax of inflectionlessness*. Philadelphia, PA: University of Pennsylvania dissertation.

Aelbrecht, Lobke. 2010. *The syntactic licensing of ellipsis*. Amsterdam: John Benjamins.

Alexiadou, Artemis. 2017. Gender and nominal ellipsis. In Nicholas LaCara, Keir Moulton & Anne-Michelle Tessier (eds.), *A schrift to fest Kyle Johnson,* 11–22. Linguistics Open Access Publications.

Arregi, Karlos & Andrew Nevins. 2014. A monoradical approach to some cases of disuppletion. *Theoretical Linguistics* 40(3/4). 311–330. https://doi.org/10.1515/tl-2014-0014.

Barrie, Michael. 2016. Gender mismatches and ellipsis in Cayuga. *Journal of Cognitive Science* 17(3). 361–388.

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

Boeckx, Cedric & Howard Lasnik. 2006. Intervention and repair. *Linguistic Inquiry* 37 (1). 150–155. https://doi.org/10.1162/ling.2006.37.1.150.

Borer, Hagit. 2005. *Structuring sense. Vol. I: In name only*. Oxford: Oxford University Press.

[15] Both reviewers raised an important question concerning acquisition: how do learners come to categorize noun pairs into the three noun classes? While authors such as Sudo & Spathas (2020) have identified independent patterns that distinguish the noun classes having to do with focus and pluralization (though complications arise for the asymmetric Class III, see Sudo & Spathas 2020: 32; see also Ranero 2021: 336 on the interaction of grammatical number and gender), it is not clear that these patterns would be available and informative during the acquisition process. As noted by one reviewer, we should strive to identify an independent empirical domain that is *observable* (see Yang 2001) and results in the featural representation that leads to the three-way contrast between nominal pairs as a consequence – assuming our proposed identity condition. Perhaps identifying variable exposure to such data would also shed light on the range of microvariation along multiple dimensions that we identified in the conclusion.

Brucart, Josep María & Jonathan E. MacDonald. 2012. Empty categories and ellipsis. In José Ignacio Hualde, Antxon Olarrea & Erin O'Rourke (eds.), *The handbook of Hispanic linguistics,* 579–601. Hoboken, NJ: Blackwell Publishing.
Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
Chung, Sandra. 2006. Sluicing and the lexicon: The point of no return. *Proceedings of BLS* 31. 73–91. https://doi.org/10.3765/bls.v31i1.896.
Corbett, Greville. 1991. *Gender*. Cambridge: Cambridge University Press.
van Craenenbroeck, Jeroen & Jason Merchant. 2013. Ellipsis phenomena. In Marcel den Dikken (ed.), *The Cambridge handbook of generative syntax*, 701–45. Cambridge: Cambridge University Press.
Creemers, Ava, Jan Don, & Paula Fenger. 2018. Some affixes are roots, others are heads. *Natural Language and Linguistic Theory* 36. 45–84. https://doi.org/10.1007/s11049-017-9372-1.
Depiante, Marcela. 2004. Dos casos de elipsis con partícula de polaridad en español. *RASAL* 1. 53–69.
Donatelli, Lucia Elizabeth. 2019. *The morphosemantics of Spanish gender: Evidence from small nominals*. Washington, DC: Georgetown University dissertation.
Embick, David & Rolf Noyer. 2007. Distributed morphology and the syntax/morphology interface. In Gillian Ramchand & Charles Reiss (eds.), *The Oxford handbook of linguistic interfaces,* 289–324. Oxford: Oxford University Press.
Halle, Morris & Alec Marantz. 1993. Distributed Morphology and the pieces of inflection. In Kenneth Hale and Samuel Jay Keyser (eds.), *The view from Building 20: Essays in linguistics in honor of Sylvain Bromberger,* 111–176. Cambridge, MA: MIT Press.
Hankamer, Jorge & Ivan Sag. 1976. Deep and surface anaphora. *Linguistic Inquiry* 7(3). 391–428.
Harley, Heidi. 2014. On the identity of roots. *Theoretical Linguistics* 40 (3). 225–276. https://doi.org/10.1515/tl-2014-0010.
Harley, Heidi & Rolf Noyer. 1999. Distributed morphology (state-of-the-article). *Glot International* 4: 3–9.
Harris, James. 1991. The exponence of gender in Spanish. *Linguistic Inquiry* 22. 27–62.
Kramer, Ruth. 2015. *The morphosyntax of gender*. Oxford: Oxford University Press.
Kramer, Ruth. 2020. Grammatical gender: A Close look at gender assignment across languages. *Annual Review of Linguistics* 6. 45–66. https://doi.org/10.1146/annurev-linguistics-011718-012450.
Lasnik, Howard. 2001. When can you save a structure by destroying it? *NELS* 31. 301–320.
Lipták, Anikó. 2013. Identity in ellipsis: An introduction. *Lingua* 166. 155–171. https://doi.org/10.1016/j.lingua.2015.08.003.
Lowenstamm, Jean. 2015. Derivational affixes as roots, no exponence: Phasal spellout meets English stress-shift. In Artemis Alexiadou, Hagit Borer, & Florian Schäfer (eds.), *The syntax of roots and the roots of syntax*. Oxford: Oxford University Press.
Mathieu, Éric, Myriam Dali & Gita Zareikar (eds.). 2018. *Gender and noun classification*. Oxford: Oxford University Press.
Mendes, Gesoel. 2020. *Investigations on salvation and non-salvation by deletion*. College Park, MD: University of Maryland, College Park dissertation.
Mendes, Gesoel & Andrew Nevins. 2022. When ellipsis can save defectiveness and when it can't. *Linguistic Inquiry* 54 (1). 182–196. https://doi.org/10.1162/ling_a_00428.

Mendes, Gesoel & Jason Kandybowicz. 2023. Salvation by deletion in Nupe. *Linguistic Inquiry* 54 (2). 299–325. https://doi.org/10.1162/ling_a_00434.
Merchant, Jason. 2001. *The syntax of silence: Sluicing, islands and the theory of ellipsis*. Oxford: Oxford University Press.
Merchant, Jason. 2013. Voice and ellipsis. *Linguistic Inquiry* 44 (1). 77–108. https://doi.org/10.1162/LING_a_00120.
Merchant, Jason. 2014. Gender mismatches under nominal ellipsis. *Lingua* 151. 9–32. http://dx.doi.org/10.1016/j.lingua.2014.01.008.
Merchant, Jason. 2019. Ellipsis: A survey of analytical approaches. In Jeroen van Craenenbroeck & Tanja Temmerman (eds.), *The Oxford handbook of ellipsis,* 19–45. Oxford: Oxford University Press.
Murphy, Andrew, Zorica Puškar & Matías Guzmán Naranjo. 2018. Gender encoding on hybrid nouns in Bosnian/Croatian/Serbian: Experimental evidence from ellipsis. In Denisa Lenertová, Roland Meyer, Radek Šimík & Luka Szucsich (eds.), *Advances in formal Slavic linguistics 2016,* 313–336. Berlin: Language Science Press.
Nakao, Chizuru. 2009. *Island repair and non-repair by PF strategies*. College Park, MD: University of Maryland, College Park dissertation.
Norris, Mark. 2014. *A theory of nominal concord*. Santa Cruz, CA: University of California at Santa Cruz dissertation.
Nunes, Jairo & Cynthia Zocca. 2009. Lack of morphological identity and ellipsis resolution in Brazilian Portuguese. In Jairo Nunes (ed.), *Minimalist essays on Brazilian Portuguese syntax,* 215–236. Amsterdam: John Benjamins.
Picallo, M. Carme. 2017. A note on the locus and function of formal gender. *Borealis: An International Journal of Hispanic Linguistics* 6(1). 1–20. https://doi.org/10.7557/1.6.1.4097.
Polinsky, Maria. 2016. Agreement in Archi from a minimalist perspective. In Oliver Bond, Greville Corbett, Marina Chumakina, & Dunstan Broen (eds.), *Archi: Complexities of agreement in a cross-theoretical perspective*. Oxford: Oxford University Press.
Polinsky, Maria. 2020. Russian nominal ellipsis: Gender exponents and gender (mis)matches. Slides from talk given at the *15th Annual Meeting of the Slavic Linguistics Society*. Indiana University.
Preminger, Omer. 2020. Functional structure in the noun phrase: Revisiting Hebrew nominals. *Glossa: a journal of general linguistics* 5(1). 68. http://doi.org/10.5334/gjgl.1244.
Ranero, Rodrigo. 2021. *Identity conditions on ellipsis*. College Park, MD: University of Maryland, College Park dissertation.
Ranero, Rodrigo. 2023. Voice mismatches in Kaqchikel Mayan sluicing. *Proceedings of the Workshop on the Structure and Constituency of the Languages of the Americas* 24. 88–102.
Ranero, Rodrigo & Justin Royer. To appear. Deep and surface anaphora: A Mayan reappraisal. *NELS* 54.
Ritter, Elizabeth. 1991. Two functional categories in noun phrases: Evidence from Modern Hebrew. In Susan D. Rothstein (ed.), *Perspectives on phrase structure: Heads and licensing,* 37–62. San Diego, CA: Academic Press.
Rooth, Mats. 1992. Ellipsis redundancy and reduction redundancy. In Steve Berman & Arild Hestvik (eds.), *Proceedings of the Stuttgart Ellipsis Workshop*, 1–26.
Ross, John R. 1969. Guess who? *Chicago Linguistic Society* 5. 252–286.
Rudin, Deniz. 2019. Head-based syntactic identity in sluicing. *Linguistic Inquiry* 50(2). 253–283. https://doi.org/10.1162/ling_a_00308.

Saab, Andrés. 2004. *El dominio de la elipsis nominal en Español: Identidad estricta e inserción tardía*. Buenos Aires, Argentina: Universidad Nacional del Comahue Master's thesis.
Saab, Andrés. 2008. *Hacia una teoría de la identidad parcial en la ellipsis*. Buenos Aires, Argentina: Universidad de Buenos Aires dissertation.
Saab, Andrés. 2010a. (Im)possible deletions in the Spanish DP. *Iberia: An International Journal of Theoretical Linguistics* 2(2). 45–83.
Saab, Andrés. 2010b. Spanish TP-ellipsis and the theory of island repair. *Probus* 22(1): 73–116. https://doi.org/10.1515/prbs.2010.003.
Saab, Andrés. 2019. Nominal ellipsis. In Jeroen van Craenenbroeck & Tanja Temmerman (eds.), *The Oxford handbook of ellipsis,* 526–561. Oxford: Oxford University Press.
Siddiqi, Daniel. 2009. *Syntax within the word: Economy, allomorphy, and argument selection in Distributed Morphology*. Amsterdam: John Benjamins.
Stockwell, Richard & Deborah J. M. Wong. 2020. Sprouting and the structure of except-phrases. *NELS* 50. 169–182.
Spathas, Giorgos & Yasutada Sudo. 2020. Gender on animal nouns in Greek. *Catalan Journal of Linguistics* 19. 25–48. https://doi.org/10.5565/rev/catjl.297.
Sudo, Yasutada & Giorgos Spathas. 2016. *Natural gender and interpretation in Greek*. Manuscript: University College London and Humboldt-Universität zu Berlin/Universität Stuttgart.
Sudo, Yasutada & Giorgos Spathas. 2020. Gender and interpretation in Greek: Comments on Merchant (2014). *Glossa: a journal of general linguistic*s 5(1): 129. https://doi.org/10.5334/gjgl.1173.
Yang, Charles. 2001. Internal and external forces in language change. *Language Variation and Change* 12. 231–250. https://doi.org/10.1017/S0954394500123014.