ABSTRACT

Title of Dissertation:	IDENTITY CONDITIONS ON ELLIPSIS
	Rodrigo Ranero Echeverría, Doctor of Philosophy, 2021
Dissertation directed by:	Professor Maria Polinsky Department of Linguistics
	Associate Professor Omer Preminger Department of Linguistics

This dissertation presents a new perspective on the identity condition underpinning ellipsis in natural language. It argues that the condition is irreducibly syntactic—at least in part—but the way this syntactic component works is different than previously thought. First, instead of simple identity of structures/features, the condition relies on non-distinctness. For example, a privative feature present in the antecedent but not in the ellipsis site (or vice-versa) does not constitute a violation of identity. Nor does a functional projection present in one but not the other. Second, the identity condition includes a component that pertains to √ROOTs. Unlike the component requiring featural non-distinctness, √ROOTs in the ellipsis site and the antecedent must be strictly identical.

After providing an overview of the core research questions surrounding ellipsis, the dissertation builds its initial case in chapter 2 on the basis of novel data from Kaqchikel (Mayan). In contrast to the pattern familiar from languages like English, Kaqchikel

bans certain voice mismatches under sluicing, but allows others. To account for that, I argue that clauses in the Agent Focus voice—which can mismatch with active and passive clauses—lack the VoiceP layer. The proposed identity condition which relies on non-distinctness captures this newly-established pattern. The empirical scope is expanded in chapter 3, where I consider mismatches above VoiceP in several languages. I show that the proposed identity condition can account for the observed generalizations regarding tense, polarity, illocution, and modality mismatches, which remain unexplained under other proposals. Chapter 4 zooms into the nominal domain and discusses mismatches in grammatical gender under nominal ellipsis in argument and predicate positions. I present cross-linguistically recurrent patterns of well-formed and ill-formed mismatches and argue that the proposed identity condition (coupled with the independently motivated mechanism of repair-by-ellipsis of morphophonological gaps) is necessary and sufficient to account for the attested patterns. I also argue that certain configurations satisfy the identity condition but are ill-formed for other reasons; in particular, ellipsis cannot repair encyclopedic gaps. Extensions of the proposal are discussed in chapter 5, including voice mismatches under sluicing in Austronesian languages, Chung's generalization, and vehicle change phenomena.

IDENTITY CONDITIONS ON ELLIPSIS

by

Rodrigo Ranero Echeverría

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2021

Advisory Committee:
 Professor Maria Polinsky, Chair
 Associate Professor Omer Preminger, Chair
 Distinguished University Professor Howard Lasnik
 Professor Norbert Hornstein
 Associate Professor Michael Israel, Dean's Representative

Dedication

Esta tesis se la dedico a mi mamá, Mely—no hay palabras suficientes para expresarte las gracias.

Acknowledgements

This dissertation would not exist if I had not struck up a conversation at a party in 2009 with Carlos Salvadó Moore, a retired mathematician and dear friend of my brother's. I told Carlitos—a tall, imposing man, reminiscent of a wizard, who sported a long white beard, a hat, and a cane—about my intention of applying to US colleges. He suggested that I look into studying linguistics (a discipline then unknown to me). I took that advice to heart and twelve years later, you're now perusing this long document on the syntax of ellipsis. Thank you, Carlitos, for your sage advice, wherever you might be—hopefully, somewhere that resembles your beloved San Juan la Laguna.

I have many other people to thank, so this will be long. I'll start with my committee members.

Thank you to Masha Polinsky and Omer Preminger—a formidable advising team. M & O have supported me from the first day I arrived at UMD. Actually, they began to mentor me *before* I set foot in Maryland, when they flew down to Guatemala during the summer of 2016 for a first, and very memorable, trip to Tecpán and Patzún. They met with me on a weekly basis for five years, entertaining my half-baked ideas, providing me with thorough feedback, and treating me like a colleague. Outside of work, they provided the right words and advice when the going got tough (and it sure did), reminding me that the most important thing was to take care of myself. I hope we get to celebrate soon (and in person) the end of this PhD journey.

Thank you to Howard Lasnik, whose office door was always open and who introduced me to the world of ellipsis. I am grateful that he agreed to be on my committee, raising

tough questions along the way that significantly improved the argumentation here. He has taught me a lot about syntax and perhaps even more about what good teaching looks like. I thank him for all the lessons and wisdom he imparted over these past 5 years.

Thank you to Norbert Hornstein, who in the before days would frequently pop into my office unannounced, ask me what I was working on, provide excellent feedback, and then tell me about the most recent paper he'd found interesting, only to head upstairs immediately after to chat with another student. I miss these conversations, but was very happy that Norbert could be on my committee, and I appreciate the enthusiasm he has shown for my work.

Finally, thank you to Michael Israel for agreeing without any qualms to be the Dean's Representative on a very tight schedule—the feedback he provided during the defense was very helpful and improved the presentation significantly.

Beyond the members of my committee, many others have made my five years in Maryland a memorable experience. First, thank you to Kim Kwok, who is a superstar (no, really – just go and read the acknowledgments sections in other dissertations). There isn't a single problem that Kim can't solve expediently; I don't know what I would have done without her! Thank you as well to the many faculty members who contributed to a great five years: Juan Uriagereka, whose friendship and mentorship helped me through some tough times; Peggy Antonisse and Tonia Bleam, who I loved TAing for and supported me in everything, inside and outside of the classroom; Alexander Williams and Valentine Hacquard, semanticists and conversationalists extraordinaire; and all others whose chats in the lunch room, during colloquia, practice talks, and regular office days made my experience so much richer.

en Chiquimulilla, gracias a todos los miembros del COPXIG, especialmente a Juan Pablo López, Felipe de la Cruz, Ramiro López, y Don Demetrio López (QEPD).

Por último, tengo que mencionar a mi querida familia, aunque me es difícil expresar solamente con un par de palabras todo lo que quisiera decir. Gracias a mi hermano, Andrés—fue un gusto que trabajáramos juntos en el 2017 y espero que pronto nos volvamos a ver (Australia está demasiado lejos). Mis papás, a quienes adoro, han sido extraordinarios, apoyándome incondicionalmente en todo momento, especialmente cuando la cosa se puso color de hormiga. Gracias a ambos—Enrique y Mely—por su amor, su trabajo duro, y por asegurarse que nada me faltara a través de los años. Sin embargo, tengo que darle unas gracias especiales a mi mamá, la persona más fuerte que conozco. Sin ella, no estaría donde estoy y *nada* de esto hubiera sido posible.

Table of Contents

List of Abbreviations

A = set A (ergative/genitive) agreement
ACT = active voice
AGR = agreement
AF = Agent Focus
AP = antipassive
APO = oblique antipassive
AT = actor trigger (voice)
B = set B (absolutive) agreement
BEN = benefactive
CAUS = causative
CLF = classifier
COM = completive aspect
COMP = complementizer
CT = circumstantial trigger (voice)
COP = copula
DEM = demonstrative
DET = determiner
DIM = diminutive
DIR = directional
EMPH = emphatic particle
EXIST = existential
F = feminine
FOC = focus particle
FP = fronting particle
IMP = imperative
INC = incompletive aspect
INST = instrumental
INT = intensifier particle
IRR = irrealis particle
L = linker
M = masculine
NEG = negation
PL = plural
PART = particle
PAS = passive
PREP = preposition
PRES = present tense
PROG = progressive
PRF = perfect
Q = question
REFL = reflexive

REL = relativizer
RN = relational noun
S = singular
SS = status suffix
TRANS = transitive
TT = theme trigger (voice)
UNM = unmarked case

Chapter 1: Setting the stage

This dissertation is about *ellipsis*: structure and meaning without form. In the case of spoken languages, ellipsis is *silence* whose meaning must be recovered through a linguistic antecedent in the discourse. The key question that this dissertation seeks to answer in relation to ellipsis has to do with the relationship between what is missing and what is pronounced:

(1)   *The identity question* (Merchant 2019: 21)
        What is the relationship between the understood material in the ellipsis site and its antecedent?

The purpose of this introductory chapter is to lay out my theoretical assumptions regarding the derivation of ellipsis. By establishing these assumptions, we will be able to focus on an answer to the question above in subsequent chapters.

Let us begin by illustrating the phenomenon at the heart of the present work. Assume, for now, that there is structure in the material that is left silent (more on this below). Consider the following examples, which illustrate different elliptical constructions—*sluicing* (2)b, auxiliary-stranding *VP ellipsis* (2)d, and *NP ellipsis* in Spanish (2)f:

(2)   *Ellipsis—meaning without form necessitating an overt antecedent*
        a. Someone was crowned America's next drag superstar, but I can't remember who was crowned America's next drag superstar.
        b. Someone was crowned America's next drag superstar, but I can't remember who.
        c. Have you watched the latest season of UNHhhh? I haven't watched the latest season of UNHhhh.
        d. Have you watched the latest season of UNHhhh? I haven't.
        e. Las   películas de Varda y   las   películas de Buñuel son mis
           the.F.PL movies   of Varda and the.F.PL movies   of Buñuel are my
           favoritas.
           favorite
           'Varda's films and Buñuel's films are my favorite.'

f. Las películas de Varda y las de Buñuel son mis favoritas.
   'Varda's films and Buñuel's are my favorite.'

How do speakers understand utterances like (2)b,d,f above, which are incomplete on the surface? The bracketed portions below indicate the apparent content of the *ellipsis sites* in these examples:

(3)   *Ellipsis: what's in the silence?*
   b. Someone was crowned America's next drag superstar, but I can't remember who
      <was crowned America's next drag superstar>.
   d. Have you watched the latest season of UNHhhh? I haven't <watched the latest season of UNHhhh>.
   f. Las películas de Varda y las <películas> de Buñuel son mis favoritas.
      'Varda's films and Buñuel's are my favorite.'

Based solely on the limited data set above, a reasonable hypothesis would maintain that ellipsis sites are structurally isomorphic to their antecedent.

It is not the case, however, that all ellipsis sites match one-to-one with their antecedent. Instead, decades of research on ellipsis have established that ellipsis sites can *mismatch* (to different degrees) with the antecedent. Consider a couple of representative examples where ellipsis is well-formed, *despite* differences between the material that is pronounced and the material that is silent:

(4)   *Ellipsis: a sample of well-formed mismatches*
   a. They can't cook fideuà, they don't know how <to cook fideuà> !
                                                                tense/finiteness mismatch
   b. This problem was to have been looked into, but obviously nobody did <look into this problem>.   *voice mismatch (Kehler 2002)*
   c. Laughton directed one masterpiece, but Varda directed many <masterpieces>.
                                                           grammatical number mismatch

In contrast to the examples above, ellipsis is sometimes ill-formed, apparently *as a result* of a mismatch:

(5)   *Ellipsis: a sample of ill-formed mismatches*
        a. *Yoli tried to melt the cup, even though it wouldn't <melt>.
                                                        argument structure mismatch
        b. *Someone crowned Bob, but we don't know by who <Bob was crowned>.
                                                        voice mismatch
        c. *Nunca llegaron la   hermana de Ana y   el   <hermano> de Sebas.
           never arrived the.F sister   of Ana and the.M brother   of Sebas
           Intended: 'Ana's sister and Sebas's brother never arrived.'
                                                        grammatical gender mismatch (Spanish)

We are thus faced with a tension between ellipsis which is well-formed despite some

mismatches, like in (4), and ellipsis that fails because of a mismatch, as in (5). This kind

of asymmetry will be the empirical purview of this dissertation.

   I will show in the present work that syntax plays a key role in determining when

mismatches are well-formed or ill-formed (see 1.3 below). Using examples like those

discussed above and others that are similar in nature, I will therefore seek to answer a

narrower question than the one that opened this chapter:

(6)   *The identity question in this dissertation*
        What is the **syntactic** relationship between the understood material in the ellipsis
        site and its antecedent?

This dissertation's main contribution is a new formulation of a purely syntactic *identity*

condition on ellipsis that, I will argue, can derive the kind of asymmetries exemplified by

the data in (4) and (5).[1] I will provide substantial empirical support for a condition that

contains two statements, each specific to different syntactic primitives, namely features

and √ROOTs. The first statement specifies the *featural* relationship that must be satisfied

---

[1] As noted by van Craenenbroeck & Merchant 2013, Merchant 2013a, Merchant 2019, and others, the identity

condition has gone by many names, which are sometimes used interchangeably, but occasionally refer to

specific aspects of the relationship that must hold between the ellipsis site and antecedent: identity

condition/requirement, identification requirement, recoverability condition, parallelism requirement, etc.

between the ellipsis site and the antecedent. The second statement specifies that √ROOTs, unlike features, must be identical between the antecedent and the ellipsis site. My proposal is shown below:

(7)   *Syntactic identity condition on ellipsis (final)*
    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 first statement in (7)a harkens back to Chomsky 1965's discussion of recoverability of deletion in comparatives, where the potential relevance of the formal notion of non-distinctness was first highlighted (see Lipták 2015, Lasnik & Funakoshi 2019 for discussion; see also Ingria 1990 for a proposal that feature matching in agreement involves a non-distinctness component). The second statement follows a proposal in Saab 2008 and is justified in chapter 4 of the present work.

In the rest of this chapter, I will lay out several theoretical assumptions that will allow us to focus exclusively on the proposal in (7). Following Merchant 2019, I first discuss the three most important issues in research on ellipsis: structure, licensing, and identity. In section 1.1, I lay out my assumptions regarding the **structure question**. I delve into the strongest evidence for the position that ellipsis sites contain unpronounced structure and adopt the proposal that ellipsis involves deletion (or non-insertion) of said structure. Section 1.2 addresses the **licensing question**. I adopt the [E] feature approach to ellipsis licensing (Merchant 2001), the proposal that a head bearing said feature in the syntax licenses the ellipsis of its complement. What is crucial for the purposes of this dissertation is that all the data analyzed herein are of two kinds: (i) well-formed examples where ellipsis is licensed by the [E] feature (i.e., the structural configuration for ellipsis is met) and the

identity condition is satisfied, and (ii) ill-formed examples where ellipsis is licensed by the [E] feature but the identity condition is not satisfied. By focusing only on such cases, we can be certain that the ill-formed status of specific utterances is the result of a violation of the identity condition. Next, the **identity question** is taken up in 1.3. I summarize different approaches to the identity condition—some semantic in nature, others syntactic, and still others a hybrid of the two. I also discuss proposals that prioritize pragmatics in lieu of or in addition to an identity condition. I will establish that whatever the contribution of pragmatic reasoning is to the well-formedness of different instances of ellipsis, there is strong evidence for the existence of an identity condition in which syntax plays a key role. Moving on to section 1.4, I discuss the **methodological question** of establishing the content of an ellipsis site. I propose that the data that can adjudicate between different approaches to the identity condition are those in which the remnant is manipulated in order to force one, and (ideally) only one, possible content for an ellipsis site. I argue that other approaches that rely on the most natural fill-in for the silence (e.g., Anand et al. 2021) are untenable, and a subset of the data which such approaches rely on cannot adjudicate between different formulations of the identity condition. Finally, section 1.5 summarizes the key empirical and theoretical takeaways of each of the chapters that make up this dissertation.

Let us start, then, with the structure question.

1.1 The structure question

Merchant 2019 phrases the structure question as follows:

(8)   *The structure question* (Merchant 2019: 21)
      In elliptical constructions, is there syntactic structure that is unpronounced?

In this dissertation, I will assume that the elliptical constructions under consideration involve syntactic structure that is not externalized. Specifically, I adopt the approach that has come to be known as PF-deletion, which dates back to the seminal work of Ross 1969a (see e.g., Sag 1976, Lasnik 2001, Baltin 2012). In more recent formulations, this approach has been formalized through the suppression of insertion in the post-syntactic component (Saab 2008, 2021, Aelbrecht 2010, Murphy 2016, Park 2017, Banerjee 2020; see Wasow 1972). For a prominent alternative to the PF-deletion approach that also assumes structure in the ellipsis site—the so-called LF-copying approach—I refer the reader to Wasow 1972, Fiengo & May 1994, Chung et al. 1995, 2011 (see Merchant 2019 for discussion and comparison of these two competing approaches). A third approach is that in ellipsis, what you see is what you get (so-called WYSIWYG approaches)—i.e., there is not structure in the silence, in all or in a subset of elliptical constructions (for one example of this family of approaches, see Culicover & Jackendoff 2019; for criticisms of WYSIWYG approaches, see Merchant 2019). My own work, presented here, offers a novel argument against WYSIWYG approaches. For example, it is unclear how a proposal without structure in the ellipsis site could make sense of the asymmetric status of voice mismatches under sluicing cross-linguistically. I will show that it is necessary to make reference to structure in the silence in order to capture why this type of mismatch under slucing is ill-formed uniformly in languages like English (see for example (5)b), whereas in languages like Kaqchikel, voice mismatches are well-formed in a subset of configurations (summarized in (30) below; see chapter 2 for details).

There exists significant empirical support for the hypothesis that ellipsis involves structure that is not externalized. In what follows, I lay out the strongest evidence: (i) case-

connectivity in sluicing (1.1.1), (ii) agreement connectivity (1.1.2), and (iii) sub-extraction asymmetries (1.1.3). Section 1.1.4 summarizes the discussion and delves into the distinction between surface and deep anaphora based primarily on the requirement of a linguistic antecedent (Hankamer & Sag 1976). There are also pockets of challenges for the structural approach to ellipsis, most prominently cases where connectivity breaks down. For example, there is debate regarding whether fragment answers involve ellipsis or not (compare Merchant 2004 to Jacobson 2016 and Goldberg & Perek 2019). It would be impossible to do justice to these debates in this introductory chapter—I refer the reader to Merchant 2019 and Lasnik & Funakoshi 2019 for discussion.

It is important to note at this juncture that, throughout this dissertation, I will circle back to the kind of evidence discussed in this section whenever necessary, especially when specific elliptical constructions form the basis of a discussion. For example, in chapter 2, I lay out a diagnostic that is specific to Kaqchikel that suggests that sluicing involves regular *wh*-movement followed by deletion (or non-insertion).

1.1.1 Case connectivity

Sluicing is the name given to clausal ellipsis with a *wh*-remnant. In his discussion of languages with morphological case marking, Ross (1969)a famously noted that the *wh*-remnant in sluicing bears the case that is expected if there is a case assigner in the silence (for more discussion of sluicing, see chapter 2). Consider the examples below to illustrate this observation (adapted from Merchant 2001). In German, the verb *schmeicheln* 'to flatter' assigns dative case to its object (9)a, while the verb *loben* 'praise' assigns accusative case (9)c. In the elliptical cases (9)b,d, the *wh*-remnant bears the expected case and cannot bear any other:

(9)   *Case connectivity in sluicing* (adapted from Merchant 2001: 89-90)

    a. Sie   wissen nicht, {*wer /   *wen /   **wem**}   er schmeicheln
       they know not   who.NOM who.ACC **who.DAT** he flatter
       will.
       wants
       'They don't know who he wants to flatter.'

    b. Er will   jemandem   schmeicheln, aber sie wissen nicht,
       he wants someone.DAT flatter   but   they know not
       {*wer /   *wen /   **wem**}.
         who.NOM who.ACC **who.DAT**
       'He wants to flatter someone, but they don't know who.'

    c. Sie   wissen nicht, {*wer /   **wen** /   *wem}   er loben will.
       they know not   who.NOM **who.ACC** who.DAT he praise wants
       'They don't know who he wants to praise.'

    d. Er will   jemanden   loben, aber sie wissen nicht
       he wants someone.ACC praise but they know not
       {*wer /   **wen** /   *wem}
         who.NOM **who.ACC** who.DAT
       'He wants to praise someone, but they don't know who.'

The kind of facts shown above are not specific to German—similar observations have been replicated more broadly (Merchant 2001 already identified case connectivity under sluicing in 9 other languages: Greek, Russian, Polish, Czech, Slovene, Finnish, Hungarian, Hindi, and Basque). There are languages where apparent violations of case connectivity follow not from sluicing, but from other structures, and thus do not constitute a problem (see Mendes 2020: 13-15 for discussion of why these examples in Japanese, Uzbek, and Turkish do not, in the end, constitute a challenge to findings regarding connectivity and see van Craenenbroeck 2010a for discussion on cleft-sources in some instances of sluicing).

Setting some problematic cases aside, approaches that do not assume structure in the silence have to stipulate that the *wh*-remnant in sluicing bears the case that it bears. Under the structural approach, in contrast, this is derived directly without any recourse to stipulations that restate the facts. I thus take case connectivity as strong evidence that there is structure in the silence.

1.1.2 Agreement

Ross 1969a once again provides us with empirical observations that can be straightforwardly explained under a structural approach to ellipsis. Observe that a sluiced clause always controls singular agreement, irrespective of the number specification of the *wh*-remnant itself (data adapted from Lasnik & Funakoshi 2019):

(10)　　*The wh-remnant of a sluice controls singular agreement*
　　　　a. He's going to give us one old problem for the test, but **which problem isn't** clear.
　　　　b. He's going to give us one old problem for the test, but **which problem *aren't** clear.
　　　　c. He's going to give us some old problems for the test, but **which problems isn't** clear.
　　　　d. *He's going to give us some old problems for the test, but **which problems aren't** clear.

The fact that the *wh*-remnant of a sluice controls singular agreement regardless of its own number specification follows from the general behavior of clausal subjects, which control singular agreement (data modelled after Lasnik & Funakoshi 2019):

(11)　　a. [That Laughton directed only one film] is/*are tragic.
　　　　b. [Why he only directed one film] is/*are well-known.
　　　　c. [That season 14 will be filmed and Lady Bunny will be cast] now seems / *seem likely.

Circling back to examples (10)c-d, the following becomes clear: if there were no additional structure, the *wh*-phrase *which problems* should control plural agreement. Since we observe singular agreement instead, we must conclude that there is more structure than meets the eye:

(12)　　He's going to give us some old problems for the test, but which problems <**he's going to give us**> isn't/*aren't clear.

A related observation regarding agreement is that an element inside of a (putative) ellipsis site controls agreement on an element outside of the ellipsis site (see again Ross 1969a).

The following VP ellipsis examples from Merchant 2013b illustrate the observation (for an argument from Polish, see Mendes 2020):

(13)   *Agreement controllers in the silence*
      a. First, there were bananas available, and then there weren't.
      b. First, there were going to be bananas available, and then there weren't.

As noted by Merchant, these agreement facts are straightforwardly derived under a structural approach to ellipsis. Non-structural approaches, in contrast, must posit the agreement connectivity as an add-on stipulation, since they assume that there is no agreement controller at all in the second clauses.[2]

    As with case connectivity, the agreement facts illustrated here offer evidence that there is structure in the silence.

1.1.3 Sub-extraction asymmetries

As discussed in Merchant 2013b, it can be established that movement dependencies of different types (A'-, A-, and X^0-) are launched from (putative) ellipsis sites. Under the structural approach to ellipsis, this is no surprise—there is structure in the silence from which to launch a movement dependency.

    Consider as an illustration the asymmetric behavior of *wh*-movement in two surface-similar configurations. The first involves VP ellipsis, the latter involves Null Complement Anaphora (NCA). As shown below, *wh*-movement from the (putative) elliptical gap in VP ellipsis is possible, whereas movement from the NCA version is not (Merchant 2013b):

---

[2] Non-structural approaches to ellipsis might propose that agreement in (10) and (13) is controlled semantically. I refer the reader to Lyskawa 2021: 39-59 for arguments against a semantic approach to agreement.

(14)   *Sub-extraction asymmetries*
          a. Which films did he refuse to see and which films did he agree to see?
          b. Which films did he refuse to see and which films did he agree to?   *VPE*
          c. *Which films did he refuse to see and which films did he agree?   *NCA*

The structural approach to ellipsis can straightforwardly account for the asymmetry above. Even though both (14)b and (14)c involve missing elements, there is only structure in the gap of the former (for other asymmetries between ellipsis and NCA, see Hankamer & Sag 1976, Depiante 2001; see Aelbrecht 2010 for some qualifications). Under non-structural approaches to ellipsis, it is difficult to straightforwardly explain this asymmetry. I will return to sub-extraction in chapter 3 in the context of Spanish predicate ellipsis and chapter 4 in the context of NP ellipsis.

1.1.4 Deep and surface anaphora

I have provided three pieces of evidence for structure in ellipsis sites—(i) case connectivity effects, (ii) agreement connectivity effects, and (iii) sub-extraction asymmetries. Other diagnostics have been proposed in the literature, with some controversy regarding their applicability (e.g., the Missing Antecedent Phenomenon; Grinder & Postal 1971, Hankamer & Sag 1976; see Merchant 2013b).

Before moving on, it is necessary to lay out a distinction that remains useful today, namely, the distinction between anaphoric expressions that require a linguistic antecedent and those that do not (Hankamer & Sag 1976). Focusing on their discussion of silent expressions, Hankamer and Sag observed that some of these require a linguistic antecedent—i.e., they must be *syntactically controlled*. These were dubbed *surface anaphora*. Other anaphoric expressions (both silent and overt) can be *pragmatically controlled*—i.e., a rich enough non-linguistic context is enough to allow speakers to

recover a meaning. These were dubbed *deep anaphora*. Consider the distinction between VP-ellipsis (an example of surface anaphora) and NCA (an example of deep anaphora) in this respect:[3]

(15)   *VP ellipsis requires syntactic control (Hankamer and Sag 1976)*
        Context: Sag produces a cleaver and prepares to hack off his left hand.
        Hankamer: #Don't be alarmed, ladies and gentlemen, we've rehearsed this act
                  several times and he never actually does.

(16)   *VP ellipsis requires syntactic control*
        I know that it looks like he is about to cut his hand, but he never actually does <cut
        his hand>.

(17)   *NCA does not require syntactic control; (Hankamer & Sag 1976)*
        Context: Two people are disturbed by loud noises of popcorn-eating in adjacent
                   row.
        Don't you think we should complain?

Hankamer & Sag correlated the requirement of syntactic control with other diagnostics that, taken together, suggested that surface anaphora involve structure, whereas deep anaphora do not.[4]

The observation that not all silence is created equal remains relevant today and will play a role in our discussion moving forward. Specifically, at different points I will assess whether the "elliptical" constructions under analysis involve true ellipsis—i.e., I will discuss whether they are surface anaphora, to use Hankamer & Sag's terminology. Only if

---

[3] There are cases where VP ellipsis seems to be allowed in the absence of syntactic control. I refer the reader to Hankamer 1978 for a convincing argument that the existence of these examples should not lead us to conclude that VP ellipsis altogether does not involve structure (see comments in Merchant 2013b).

[4] Some have taken cases like (17) to involve an intransitive use of the relevant verbs (e.g., Napoli 1983). For arguments against such an analysis, see Depiante 2019.

we determine that this is the case can we reasonably conclude that our analysis of such data bears on the identity condition on ellipsis, and not some other area of the grammar.

1.2 The licensing question

Merchant 2019 lays out the licensing question as follows:

(18) *The licensing question* (Merchant 2019: 22)
      What heads or positions or structures allow for 'ellipsis' and what are the locality conditions on the relation between these structures and ellipsis?

This question is raised as a result of the following truism: ellipsis is not allowed in just *any* structural configuration, even if a fully identical antecedent is available. Consider the following data (adapted from van Craenenbroeck & Merchant 2013):

(19) *Ellipsis requires more than just an identical antecedent*
      a. *Moby Dick was being discussed and War and Peace was being <discussed> too.
      b. Moby Dick was discussed and War and Peace was <discussed> too.

Other cases illustrate the same observation, perhaps more strikingly. For instance, VP ellipsis with a *to* remnant is allowed in some, but not all configurations (see Lobeck 1995, Martin 2001). Once again, the question of there being an identical antecedent is not relevant in accounting for the ill-formed status of examples like (20)d,h below, since such an antecedent *is* available:

(20) *Licensing* (adapted from Zwicky 1981 cited by Lobeck 1995: 165)
      a. Even though he doesn't like to jog every day, Ron jogs every day.
      b. Even though he doesn't like to <jog every day>, Ron jogs every day.
      c. Even though he could jog to stay in shape, Ron doesn't do anything to stay in shape.
      d. *Even though he could jog to <stay in shape>, Ron doesn't do anything to stay in shape.
      e. You shouldn't play with rifles, because it's dangerous to play with rifles.
      f. You shouldn't play with rifles, because it's dangerous to <play with rifles>.
      g. You shouldn't play with rifles, because to play with rifles is dangerous.
      h. *You shouldn't play with rifles, because to <play with rifles> is dangerous.

The same kind of observation can be illustrated by comparing the availability of different

elliptical constructions across languages. For example, whereas English allows auxiliary-

stranding VP ellipsis, languages like Spanish do not:

(21)   *VP ellipsis in English, not Spanish; stranded auxiliary 'have'*
      a. Amanda has written on the film, and Sean has written on the film too.
      b. Amanda ha escrito sobre la película, y   Sean ha escrito sobre
         Amanda has written on   the film   and Sean has written on
         la película también.
         the film   too
      c. Amanda has written on the film, and Sean has <written on the film> too.
      d. *Amanda ha escrito sobre la película, y Sean ha <escrito sobre la película>
         también

(22)   *VP ellipsis in English, not Spanish; stranded auxiliary 'be'*
      a. Devika is talking about the film, and Nick is talking about the film too.
      b. Devika está hablando sobre la película y   Nick está hablando sobre
         Devika is   talking   about the film   and Nick is   talking   about
         la película también.
         the film   too
      c. Devika is talking about the film, and Nick is <talking about the film> too.
      d. *Devika está hablando sobre la película y Nick está <hablando sobre la película>
         también.

In a nutshell, it is simply not the case that structure can be omitted freely as long as an

identical antecedent is available. Instead, there must be a way to capture the fact that

ellipsis, even when an identity violation is not in question, is allowed only in a subset of

structural configurations.

In this dissertation, I will adopt the viewpoint that ellipsis is licensed by a head bearing

an [E] feature (Merchant 2001). I will assume that the [E] feature is compatible only with

specific heads on a language-particular basis. For example, the difference between English

and Spanish regarding auxiliary stranding VP ellipsis is a lexical matter: if VP ellipsis

involves the deletion of *v*P (which is the complement of Voice; see chapter 2), we can

encode the difference among these languages by positing that Voice in English can bear the [E] feature, whereas Voice is Spanish cannot (notice that capturing cross-linguistic variation in this manner is in line with Borer (1984)'s conjecture that parametric differences can be attributed to the presence or absence of features on particular heads). I will also assume that the [E] feature licenses ellipsis and ellipsis is well-formed only if the identity condition—a *general* condition in the grammar that is universal and not construction specific (see chapter 3)—is satisfied.[5]

(23)   *The [E] feature*
      a. Is compatible with a subset of heads in the lexicon.
      b. Marks its complement for deletion (or non-insertion) in the post-syntactic component of the grammar.

The above will be enough for the purposes of this dissertation; I refer the reader to van Craenenbroeck & Merchant 2013 and Merchant 2019 for discussion on more specific aspects of the [E]-feature; see also Aelbrecht 2010 for long-distance licensing mediated by Agree and Conner 2019 for a requirement that licensing heads be overt in some instances.[6]

In all the examples that will be analyzed to adjudicate between the proposal in this dissertation and other approaches, the question of whether the [E] feature is borne by the relevant head will not be at stake. In other words, it will be possible to attribute the ill-formed status of specific manipulations solely to a violation of the identity condition. I now turn to this final issue.

---

[5] See Merchant 2001, van Craenenbroeck & Merchant 2013: 729, 744 (endnote 32) for a proposal that encodes the identification/recoverability requirement on the elided phrase (Merchant 2001's e-GIVENness) as the semantics of the [E] feature.

[6] For alternatives to the [E] feature approach to licensing, see Saito & Murasugi 1990, Lobeck 1995, and Thoms 2010; see also Zagona 1988, Johnson 2001.

1.3 The identity question

The identity question, repeated below, has attracted a great deal of attention in the literature.

(24)　*The identity question* (Merchant 2019: 21)
　　　What is the relationship between the understood material in ellipsis and its antecedent?

Here is a sample of approaches that have been put forth in the literature to answer (24). I have expanded Merchant (2019: 22's) summary of approaches by adding Chung 2006, Saab 2010a, Tanaka 2011, Murphy 2016, and Rudin 2019:

(25)　*A sample of approaches to the identity condition* (modified from Merchant 2019)
　　　a. *The relationship is semantic*: Sag & Hankamer 1984, Merchant 2001, Aelbrecht 2010, van Craenenbroeck 2010b
　　　b. *The relationship is syntactic*: Sag 1976, Williams 1977, Fiengo & May 1994, Chung et al. 1995, Fox 2000, Fox & Lasnik 2003, Saab 2010a, Tanaka 2011, Murphy 2016, Rudin 2019
　　　c. *The relationship is a hybrid of syntactic and semantic identity*: Kehler 2002, Chung 2006, Chung 2013a, Merchant 2013a.

In a nutshell, for approaches that assume that there is structure in the ellipsis site, some posit that the relationship between the ellipsis site and the antecedent is semantic in nature (25)a; others have proposed that the relation is syntactic only (25)b; most recently, hybrid approaches which incorporate a syntactic and semantic component have also been defended (25)c.[7]

　　A question may arise whether an identity condition—especially a syntactic one—is needed at all. A final family of approaches to ellipsis becomes relevant in assessing this question, given that their empirical coverage may overlap with the work done by an identity

---

[7] Kehler (2002) also proposes pragmatic conditions on discourse coherence within his approach; see Kehler 2019 for discussion.

condition, however it is formulated. These approaches concentrate on pragmatic considerations that govern the acceptability of different elliptical examples (see e.g., Kertz 2008, 2013). The account of voice mismatches in VP ellipsis is useful to represent the analytical flavor of this type of approach. Consider, for example, the contrast between the following examples, where the same kind of voice mismatch exists (passive in the first clause, active in the second). Nevertheless, only the first example is judged as well-formed:

(26)   *A contrast in voice mismatch in VP ellipsis* (Kehler 2019)
    a. #This problem was looked into **by John**, and **Bob** did too.   *Passive-Active
    b. This problem was to have been looked into, but obviously nobody did.
                                                                      ✓Passive-Active

According to Kertz (2013), a problem related to information structure is to blame for the ill-formed status of (26)a. Specifically, Kertz posits that the speaker's choice of passive in the antecedent results in the Agent (*John*) being demoted to a non-topical position. This creates an issue when the second clause is considered, since the non-demoted Agent (*Bob*) functions as a contrastive topic. This asymmetry in the information-structural status of the Agents in the two clauses creates an infelicitous discourse, since the hearer is meant to backtrack and interpret *John* in the first clause as a topic—an unlikely interpretation for by-phrases in passives. In contrast, no such problem of this nature exists in (26)b, where the Agent is not mentioned in the first clause, so the voice mismatch is tolerated. However, the experimental results in Kertz 2008, 2013 show that the contrast between examples like (26)a-b is maintained even in *non-elliptical configurations*. This suggests that the relevance of the information structure status of different elements in both clauses is *independent* of ellipsis and any identity condition.

In chapter 2, I will discuss Merchant 2013a's analysis of well-formed voice mismatches under VP ellipsis like (26)b, specifically in contrast to the ill-formed status of

voice mismatches in English sluicing (see (29) below). For Merchant's analysis, the ill-formed nature of cases like (26)a would have no bearing on the identity condition on ellipsis, given that even non-elliptical controls are degraded. In other words, a proposal like Kertz 2013's for the source of the contrast between (26)a-b does not cast doubt on the necessity of a syntactico-semantic identity condition in the account of other patterns (e.g., the sluicing data discussed by Merchant and others, as well as the novel data discussed throughout the present work).

One domain where an approach based on pragmatics cannot provide insight (and positing an identity condition seems necessary) involves the availability/non-availability of preposition stranding (P-stranding) in sluicing. First, observe that examples (27)a-b show that English allows for P-stranding under *wh*-movement, in non-elliptical and sluicing examples:

(27)   *English is a P-stranding language*
        a. Trinity is jealous **of someone**, but I don't know **who** Trinity is jealous **of**.
        b. Trinity is jealous **of someone**, but I don't know **who** <Trinity is jealous **of**>.

However, P-stranding is impossible in the ellipsis site when the *wh*-remnant does not have a correlate in the antecedent, the type of sluicing known as *sprouting* (Chung et al. 1995). Observe that (28)a is a non-elliptical control showing that a preposition can be stranded in the second clause even if the *wh*-phrase has no overt correlate in the antecedent (in contrast to (27)b above, where the *wh*-phrase in the second clause has a correlate *of someone* in the first clause). The crucial example for our purposes is (28)b, which shows that a language like English—which allows P-stranding in general—does *not* allow P-stranding under sprouting.

(28)   *No P-stranding under sprouting (see Chung 2006)*
        a. Trinity is jealous, but I don't know **who** Trinity is jealous **of**.

18

b. *Trinity is jealous, but I don't know **who** <Trinity is jealous **of**>.

The facts above are strong evidence for the existence of an identity condition with a syntactic component that regulates the availability of ellipsis—an alternative explanation that seeks to derive the facts above exclusively through discourse considerations is hard to envision. It is important to note that there does not seem to be an issue of recoverability or indeterminacy as it pertains to the elided preposition *of* in (28)b: by holding constant the choice of adjective *jealous*, and the sprouting of a Theme argument, the choice of preposition is uniquely determined by the choice of adjective lexeme. I will return to P-stranding and sluicing in chapter 2 and will delve in chapter 5 into data like (28)d, which is a subcase of what has come to be known as *Chung's Generalization* (Chung 2006). Specifically, I will argue in 5.2 that Chung's Generalization follows from the √ROOT identity requirement of the present proposal (7)b.

The novel empirical results in this dissertation also argue strongly against approaches that would seek to focus only on discourse coherence (of whatever type) as the regulator of well- and ill-formed elliptical examples. Returning to voice, it has been acknowledged before that the impossibility of voice mismatches in English sluicing is a challenge for exclusively pragmatic accounts (see Kertz 2013: 422-423):

(29)   *Voice mismatches ill-formed in English sluicing*
         a. *Jaida was crowned, but we don't know who <crowned Jaida>. *Passive-Active
         b. *Someone crowned Jaida, but we don't know who by / by who <Jaida was
            crowned>.   *Active-Passive

Moving beyond English, the picture becomes more complicated for the pragmatic approaches. As I will show in chapter 2, a subset of voice mismatches in Kaqchikel Mayan—a language with a much richer voice system than English—is actually *well-*formed under sluicing:

(30)   *Voice mismatches in Kaqchikel sluicing (chapter 2 of the present work)*
      a. *Antipassive-Active
      b. ✓Active-Agent Focus
      c. ✓Agent Focus-Active
      d. ✓Passive-Agent Focus

I will demonstrate that the identity condition on ellipsis proposed here (which requires that featural non-distinctness be satisfied; see (7)a) derives the facts above, as well as the ill-formed status of voice mismatch in English sluicing (29). In contrast, it seems difficult to formulate an explanation where discourse coherence or information structure considerations derive the empirical picture of voice mismatch in Kaqchikel and English sluicing. Similarly, we will observe in chapter 3 that an asymmetry in the availability of tense mismatches between Spanish and English can be explained by the proposal advanced here, while approaches that deny the relevance of a requirement on the syntactic relationship between the ellipsis site and the antecedent would be challenging to maintain in light of those facts.

Circling back to the issue of the identity condition, I do not plan to develop an in-depth comparison of the entire range of approaches that have been taken through the years—such an endeavor would take us too far afield. Rather, I am going to take as a given that there is sufficient evidence that syntax must be, at the very least, a component in the identity condition. The main empirical results discussed in the literature that lead me adopt this assumption and build on it are (i) the asymmetry between sluicing and VP ellipsis regarding the availability of voice mismatches in English (discussed in detail in chapter 2); (ii) the ill-formed status of argument structure mismatches (see (32) below); and (iii) Chung's generalization (see chapter 5). The results in this dissertation will constitute additional evidence for such a position.

Given that syntax must play a role in the identity condition, the purpose of this dissertation is to argue for a specific formulation of this syntactic component of the identity condition (namely (7) above). At different points throughout the rest of this work, I will compare my approach to competing approaches that propose a different requirement on the syntactic relationship between the ellipsis site and the antecedent (e.g., Saab 2010a, Merchant 2013a, Rudin 2019). I will conclude that the proposal here is on the right track and brings us closer to determining this key component of the identity condition.

1.4. A methodological question

One final question must be addressed before we move on to the core chapters of the present work:

(31)   *A methodological question*
        How do we establish what is in the ellipsis site?

The question above seems trivial but is particularly crucial when our goal is to adjudicate between different versions of the identity condition.

In this dissertation, I will argue that the best way to establish the content of an ellipsis site is by manipulating the remnant surrounding the silence, such that one, and (ideally) only one possible structure underlies the silence.[8] For instance, by having a *wh*-remnant encased in a prepositional structure of one kind or another, an argument structure mismatch

---

[8] I am setting aside here the issue of evaluating whether the ellipsis site actually contains an otherwise illicit structure that is (potentially) repaired by ellipsis, depending on the nature of the violation (see chapter 2 for a discussion of island repair by ellipsis and chapter 4 for a discussion of repair asymmetries under ellipsis more generally).

is forced in the ill-formed sluicing examples below, adapted from Merchant 2013a (see chapters 2 and 3 for argument structure mismatches):

(32)　　*Forcing a mismatch via the wh-remnant*
　　　　a. They embroidered something with peace signs.
　　　　b. They embroidered peace signs on something.
　　　　c. *They embroidered something with peace signs, but I don't know what on <they embroidered peace signs>.
　　　　d. *They embroidered something on their jackets, but I don't know with what <they embroidered their jackets>. 　　(*On image impression reading of 'with what'*)

To illustrate this approach more explicitly, consider the following examples of predicate ellipsis in Spanish with a polarity particle remnant (see chapter 3 for more discussion). Here, an adverbial remnant is manipulated to ensure that only one tense specification could be in the ellipsis site—in brief, *ayer* 'yesterday' is only compatible with past tense (33)c-d:

(33)　　*Spanish TP-ellipsis—Tense clash disallowed*
　　　　a. Mañana　　no　**iré**　　　al　　parque, pero pasado mañana　sí　**iré**
　　　　　　tomorrow　NEG　**go.1S.FUT** to.the park　　but　after　tomorrow yes **go.1S.FUT**
　　　　　　al　　　parque.
　　　　　　to.the park
　　　　　　'Tomorrow, I won't go to the park, but the day after tomorrow, I will go to the park.'
　　　　b. Mañana no **iré** al parque, pero pasado mañana sí <**iré al parque**>.
　　　　　　'Tomorrow, I won't go to the park, but the day after tomorrow, I will.'
　　　　c. Mañana　　no　**iré**　　　al　　parque, pero ayer　　　sí　**fui**
　　　　　　tomorrow　NEG　**go.1S.FUT** to.the park　　but　yesterday yes **go.1S.PST**
　　　　　　al　　　parque
　　　　　　to.the park
　　　　　　'Tomorrow, I won't go to the park, but yesterday I did go to the park.'
　　　　d. *Mañana no **iré** al parque, pero ayer sí <**fui al parque**>.
　　　　　　Intended: 'Tomorrow, I won't go to the park, but yesterday, I did.
　　　　　　　　　　　　　　　　　　　　　　　　*$T_{FUTURE} - T_{PAST}$ mismatch

The result of the manipulation above provides clear evidence that there cannot be a tense clash between the antecedent and the ellipsis site. I will argue in chapter 3 that such a configuration violates the identity condition on ellipsis proposed in the present work.

In contrast to the approach that I endorse, one could adopt an alternative way of determining the content of ellipsis sites. For instance, imagine that we asked several Spanish consultants what the best way to fill in the silence is for the following example:

(34)   Hoy no voy   al   parque, pero mañana   sí *\<ellipsis site\>*.
       today not go.1S.PRES to.the park   but tomorrow yes
       'Today, I don't go to the park, but tomorrow I do.'

Imagine, further, that all these consultants provide the following as their preferred answer, where the verb in the second clause is in the future tense:

(35)   Hoy no **voy**   al   parque, pero mañana sí **iré**   al   .
       today not **go.1S.PRES** to.the park   but tomorrow yes **go.1S.FUT** to.the
       parque
       park
       'Today, I don't go to the park, but tomorrow I will go to the park.'

If we took the above as our primary indicator of the content of the ellipsis site in (34), we might be tempted to conclude that tense *can*, in fact, clash under ellipsis. This conclusion would be at odds with our conclusion from example (33). Faced with such a state of affairs, we might also attempt to formulate an identity condition that could, somehow, accommodate both sets of facts.

However, such a way of proceeding would be wholly premature and an artifact of the methodology that was used to determine the content of the ellipsis site in (34). The issue that was overlooked by using this methodological approach is the following: the adverbial remnant *mañana* 'tomorrow' in (34) did *not*, in fact, *force* a single content for the ellipsis site. Observe that a full tense *match* between the ellipsis site and the antecedent is also

possible with such a remnant— it just happened that consultants did not offer such a fill-in as their preferred alternative:

(36) Hoy no voy   al   parque, pero mañana sí **voy**   al   parque.
      today not go.1S.PRES to.the park   but tomorrow yes **go.1S.PRES** to.the park
      'Today, I don't go to the park, but tomorrow I do go to the park.'

The prior discussion should illustrate the shortcomings of a survey methodology of this sort. In brief, the question of why consultants in this hypothetical scenario would offer (35), as opposed to (36), is arguably a question about individual psychology that has no bearing on the question at hand—determining the type of mismatches that are well- or ill-formed in ellipsis.

To summarize, we can conclude two things from this discussion. First, assuming that there is more than one possible choice for the content of the ellipsis site (even holding communicative intent constant), it is irrelevant that consultants may offer one preferred fill-in or another for the ellipsis site in a well-formed string that is under consideration. Second, data in which a remnant does not force a mismatch should not be the primary type of evidence used to adjudicate between different approaches to the identity condition.

Nevertheless, several recent works have taken the kind of approach that I have just criticized, formulating in the process identity conditions that are based (partially) on results stemming from such a methodology. Consider, for example, Anand et al. 2021, which presents the following data point involving a sluice with a *how* remnant:

(37)   *Data adapted from Anand et al. 2021*
       She tried to imagine his long, sensitive fingers cleaning and oiling a gun. She understood that he must be surviving, more or less, but she couldn't figure out how *<ellipsis site>*.

What is in the ellipsis site above? According to the authors, most consultants give as their preferred fill-in '*he was surviving*'. Additionally, a variant with a modal '*he could be*

surviving' is also offered, occasionally as a first-choice. Finally, it seems that the least preferred option is a fill-in in which there is a full match with the antecedent: '*he must be surviving*'.

The authors' interpretation of this state of affairs is as follows: mismatches whose locus is above a specific portion of the clause are freely allowed under ellipsis, and such freedom should be encoded in the identity condition. In other words, they take it as significant that the preferred fill-in is one in which there is a mismatch in tense or modality; furthermore, the authors consider that this mismatch—based, once more, on a preferred fill-in—can adjudicate between different formulations of the identity condition on ellipsis.[9]

In chapter 3, I will argue that drawing this kind of conclusion from examples like (37) is on the wrong track (see section 3.6 for a discussion of Anand et al. 2021 specifically). The only thing that should be concluded from (37) is that there is an available antecedent for the ellipsis site that does *not* result in a tense or modality mismatch and complies with the identity condition; namely, the less-preferred fill in where there is a full match:

(38)   She tried to imagine his long, sensitive fingers cleaning and oiling a gun. She understood that he must be surviving, more or less, but she couldn't figure out how **he must be surviving**.

Though methodological concerns will be peppered throughout the present work, chapter 3 will engage most explicitly with the question raised at the beginning of this subsection. I will show that a range of examples akin to (37) (mostly from Rudin 2019) comply with the identity condition proposed in this dissertation and do not provide evidence for alternatives that seek to cleave the clausal spine into a portion that enters into the calculation of the

---

[9] In fact, data like (37) are taken to be counter-evidence to the first statement of the identity condition proposed in this dissertation (7)a, which was discussed in the manuscript Ranero 2019b.

identity condition and another that does not. In doing so, I will show that my methodological approach is the only way to arrive at a clearer understanding of the identity condition. In brief, data in which a single content is forced for the ellipsis site should be the preeminent adjudicator between proposals (i.e., examples like Spanish (33)).

I have now laid out my assumptions regarding four questions that arise in relation to ellipsis: (i) the structure question, (ii) the licensing question, (iii) the identity question, and (iv) methodological issues that arise when determining what is in the silence. Against this background, let me summarize Chapters 2-5 of this dissertation (Chapter 6 summarizes my main empirical and theoretical conclusions).

1.5 Summary of this dissertation

This dissertation discusses elliptical data from an array of languages, with three core chapters focusing on well- and ill-formed mismatches whose locus is at different levels of the clausal spine.

Chapter 2 forms the backbone of the present work and discusses novel sluicing data from Kaqchikel (Mayan). Empirically, I show that a subset of voice mismatches is well-formed under sluicing in the language, in stark contrast to the empirical picture offered by languages like English. I propose that the identity condition on ellipsis must include a syntactic component that requires that the antecedent and ellipsis site be featurally non-distinct. I argue that all well-formed voice mismatches in Kaqchikel involve an antecedent or ellipsis site lacking VoiceP altogether. Specifically, I propose that clauses in the Agent Focus (AF) voice—which participates in well-formed Active-AF, AF-Active, and Passive-AF sluices—instantiate clauses where the VoiceP layer has been derivationally removed by the operation Exfoliation (Pesetsky 2021). All well-formed examples thus satisfy

26

featural non-distinctness, since they do not incur in a clash of featurally specified Voice^0 heads. I show in detail that alternative analyses of the facts that would attribute the well-formed status of certain examples to repair-by-ellipsis cannot be correct. I end by demonstrating that the proposal has welcome consequences for our understanding of AF in Kaqchikel, providing insight that goes beyond the identity condition on ellipsis.

Chapter 3 applies the proposal developed in the previous chapter to a set of data that show an apparent asymmetry in mismatch availability depending on the locus of the mismatching element: whereas mismatching features within the eventive core (VoiceP/vP) seem to result in ill-formed elliptical examples in English, mismatching features above the eventive core seem to result in well-formed ones. On the basis of such data and additional observations, the chapter is structured as a detailed comparison between the present proposal and the approach in Rudin 2019, which posits a strict syntactic identity condition that ranges only over material within the eventive core. The chapter expands the empirical landscape laid out in Rudin 2019 and demonstrates that the existence of ill-formed examples cross-linguistically, where certain featural mismatches are not tolerated despite being above the eventive core, demonstrate that a non-distinctness approach is on the right track. An array of mismatches in tense, polarity, illocution, and modality in English are reanalyzed under the present proposal and are shown to comply with featural non-distinctness. I conclude that the range of mismatches observed follow directly from the flexibility offered by the proposed identity condition involving featural non-distinctness.

Chapter 4 takes us into the nominal domain and focuses on the intricate patterning of mismatches in grammatical gender. I establish an empirical generalization that is replicated across several languages; namely, that there exist three distinct classes of nouns which

exhibit distinct patterns of behavior when it comes to mismatches in grammatical gender. For Class I noun pairs, mismatches are symmetrically ill-formed; for Class II, they are symmetrically well-formed; finally, for Class III, a mismatch is well-formed only if the masculine member of the noun pair is in the antecedent. In order to account for the attested patterns, I argue that the identity condition on ellipsis must be split into two statements. The first statement is the requirement of featural non-distinctness, which will already have been discussed in chapters 2 and 3. The other statement does not pertain to features, but to √ROOTs, requiring absolute identity among such elements. Focusing on Spanish as a case study, I lay out my assumptions about the decomposition of nominals and the variety of nominalizing heads in the language bearing [GENDER] features, and show how the final identity condition proposed in this work can derive the pattern. Additionally, I discuss how certain well-formed examples are the result of repair-by-ellipsis; in contrast, some examples satisfy the identity condition, but are nevertheless ill-formed because they violate grammatical requirements that cannot be repaired by ellipsis.

Chapter 5 discusses possible extensions for the proposal in this dissertation and points out areas that may remain challenging. I discuss (i) voice mismatches in sluicing in the Austronesian languages Chamorro and Malagasy, (ii) Chung's generalization, (iii) vehicle change phenomena, (iv) indexical switches, (v) Warner's effects, and (vi) a range of data suggesting that an asymmetric subset condition on ellipsis is warranted.

Chapter 2: Voice mismatches in Kaqchikel sluicing

In this chapter, I explore well-formed and ill-formed mismatches in ellipsis at the VoiceP/vP level, focusing primarily on sluicing in Kaqchikel (Mayan). The main empirical finding that will inform my theoretical proposal is the following: in contrast to languages like English, where voice mismatches in sluicing are ill-formed across the board, a subset of voice mismatches in Kaqchikel are well-formed. In light of this novel observation, I will propose that the syntactic identity condition on ellipsis is not predicated on strict featural identity. Instead, the condition is predicated on the satisfaction of featural non-distinctness:

(1) *Syntactic identity in ellipsis (first version; to be amended in chapter 4)*
     The antecedent and material properly contained within the ellipsis site must be featurally non-distinct.

I will show that all instances of a well-formed voice mismatch in Kaqchikel satisfy (1), and argue that either the antecedent or the ellipsis site lack VoiceP altogether. The proposal in (1) will form the backbone of the analyses in chapters 3 and 4 as well, where I delve into mismatches cross-linguistically at different levels of structure—above VoiceP in chapter 3 and inside the DP in chapter 4. Eventually, I will modify the proposal in (1) by adding a requirement that √ROOTs—a distinct syntactic primitive from features—must be strictly identical between the antecedent an ellipsis site (see 4.4).

This chapter is structured as follows. In 2.1, I first illustrate the unavailability of voice and argument structure mismatches in languages like English under sluicing, describing the solution offered in Merchant 2013a that capitalizes on an asymmetry between sluicing and VP ellipsis in terms of the size of the elided constituent. I also foreshadow the results of my empirical investigation into Kaqchikel, showing that voice *can* mismatch under sluicing in that language, in a subset of possible configurations. I move on to the empirical

description in 2.2, beginning with an introduction to Kaqchikel's morphosyntax. I then describe sluicing in the language, showing that the ellipsis site contains a full-fledged syntactic structure, as opposed to a cleft or short source. I go on to show that Kaqchikel allows for some voice mismatches, but not others. Specifically, it allows for the Agent Focus voice to mismatch with active and passive voices under sluicing, whereas antipassive cannot mismatch with active. In 2.3, I provide a novel analysis of the Agent Focus voice, proposing that an Agent Focus clause instantiates a structure where the VoiceP layer has been removed via the operation Exfoliation (Pesetsky 2021). With this analysis in hand, I propose an identity condition on ellipsis that requires that the antecedent and the ellipsis site be featurally non-distinct, accounting for the Kaqchikel facts and the broader unavailability of active-passive mismatches cross-linguistically. This identity condition forms the core analytical contribution of this dissertation and is used in chapter 3 and 4 to account for a broader range of mismatches beyond the VoiceP/vP domain. In 2.4, I show how the Exfoliation analysis sheds light on the unusual agreement properties of Agent Focus, as well as several restrictions on its use depending on the make-up of the internal argument. In section 2.5, I lay out some open puzzles stemming from our discussion of sluicing and the voice inventory in Kaqchikel. Finally, 2.6 summarizes the chapter.

2.1 Mismatches in voice and argument structure

As discussed in chapter 1, the formulation of an empirically and explanatorily adequate identity condition that regulates the availability of ellipsis has been a fertile topic of debate for decades. Is the requirement purely syntactic (e.g., Fox and Lasnik 2003), semantic (e.g., Merchant 2001), or a hybrid (e.g., Merchant 2013a, Chung 2006, 2013, Barros 2014)? This chapter addresses this question head-on by reviewing and adding to an empirical domain

which has proven informative in the debate: cases in which there is a mismatch in *voice* between the contents of the antecedent and ellipsis site.

Let us illustrate the empirical puzzle concretely. Consider the following English and Spanish sluicing examples where voice mismatches between antecedent and the ellipsis sites. Sluicing is clausal ellipsis where the remnant is a *wh*-phrase. Note that the content of the ellipsis site is forced via our choice of *wh*-remnant—i.e., a wh-PP headed by 'by' forces a passive voice within the silence. The examples show that an active antecedent cannot mismatch with a passive sluice (2)b and (2)d and vice-versa, (3)b and (3)d:

(2)   *Active-Passive mismatch in sluicing*
      a. Someone hugged Jaida, but we don't know who Jaida/she was hugged by.
      b. *Someone hugged Jaida, but we don't know by who(m)/who by <Jaida was hugged>.
      c. Alguien abrazó   a Andrea, pero no sabemos   por quién
         someone hug.3SG.PST A Andrea but   not know.1P.PRS by who
         fue   abrazada Andrea/*pro*.
         be.3S.PST hugged   Andrea
         'Someone hugged Andrea, but I don't know who Andrea/she was hugged by.'
      d. *Alguien abrazó a Andrea, pero no sabemos por quién <fue abrazada Andrea/*pro*>.
         'Someone hugged Andrea, but we don't know by who/who by.'

(3)   *Passive-Active mismatch in sluicing*
      a. Jaida was hugged, but we don't know who hugged Jaida/her.
      b. *Jaida was hugged, but we don't know who <hugged Jaida/her>.
      c. Andrea fue   abrazada, pero no sabemos   quién abrazó   a
         Andrea be.3S.PST hugged   but   not know.1P.PRS who   hug.3SG.PST A
         Andrea.
         Andrea
         'Andrea was hugged, but we don't know who hugged Andrea.'
      d. *Andrea fue abrazada, pero no sabemos quién <abrazó a Andrea>.
         Intended: 'Andrea was hugged, but we don't know who.'

The examples above show that a mismatch in voice is not tolerated under sluicing. Examples in other languages like German show the same pattern (Merchant 2013: 81-82).

31

The ban on voice mismatch arises in other ellipsis types where structure containing VoiceP is targeted (pseudogapping and stripping), as well as other constructions that arguably involve ellipsis (gapping and fragment answers); see Merchant 2013a for details.

The examples we just discussed are significant for the debate on how the identity condition should be formulated. If one adopts a purely semantic condition based on mutual entailment (e.g., Merchant 2001's e-GIVENness), it is unclear how to rule out the relevant examples in (2)-(3), since both clauses in each example are truth-conditionally equivalent. As a result, this kind of data has given renewed vigor to proposals that posit that syntactic identity regulates the availability of ellipsis. Merchant (2013)a, for example, argues that a syntactic condition like the following explains the unavailability of the voice mismatch under sluicing:[10]

(4)   *Merchant (2013)a's syntactic identity condition (Chung 2013a: 3)*
        The heads in the verbal spine of the elided constituent must be syntactically identical to the corresponding heads in the antecedent.

A condition like (4) ensures that voice mismatches are disallowed, since the $Voice^0$ heads in the antecedent and ellipsis site are not identical in the relevant examples in (2)-(3). Crucially, these heads are included in the ellipsis site in sluicing. An example showing the issue with an Active-Passive mismatch is shown below; note that $Voice_{ACT}$ and $Voice_{PASS}$ are non-identical in the antecedent and the target clause:

---

[10] Merchant never spells-out his condition explicitly in the paper in question. This is an interpretation of Merchant's proposal, found in Chung 2013's discussion of that work.

(5)   *Voice mismatches in sluicing are banned; *Active-Passive* (based on Merchant 2013: 93)

    a. *Someone hugged Jaida, but we don't know by whom.

    b. *Antecedent clause*

```
              TP
         ┌─────┴─────┐
    someone₁         │
              ┌───────┴───────┐
              T             VoiceP
                      ┌───────┴───────┐
                  [Voice_ACT]        vP
                              ┌───────┴───────┐
                             t₁              │
                                    ┌─────────┴─────────┐
                                  v_trans              VP
                                              ┌─────────┴─────────┐
                                            hug               Jaida
```

    c. *Target clause*

```
                CP
         ┌───────┴───────┐
       PP₁              │
      ╱────╲     ┌───────┴───────┐
   by  whom   C_[E]          <TP_E>
                       ┌───────┴───────┐
                    Jaida₂            │
                             ┌─────────┴─────────┐
                            was                vP
                                      ┌─────────┴─────────┐
                                    t_was             VoiceP
                                             ┌──────────┴──────────┐
                                        [Voice_PASS]              vP
                                                         ┌─────────┴─────────┐
                                                        t₁                  │
                                                              ┌──────────────┴──────────────┐
                                                            v_trans                        VP
                                                                             ┌──────────────┴──────────────┐
                                                                           hug                            t₂
```
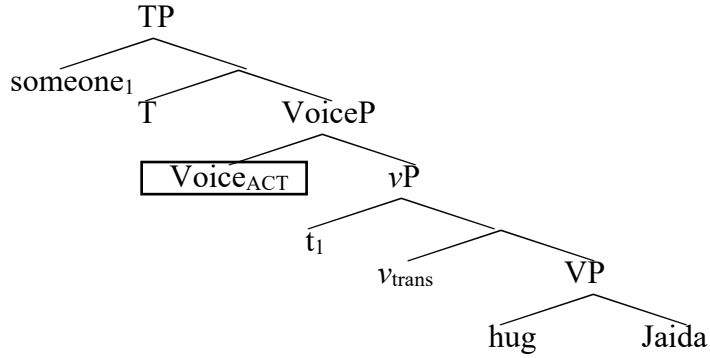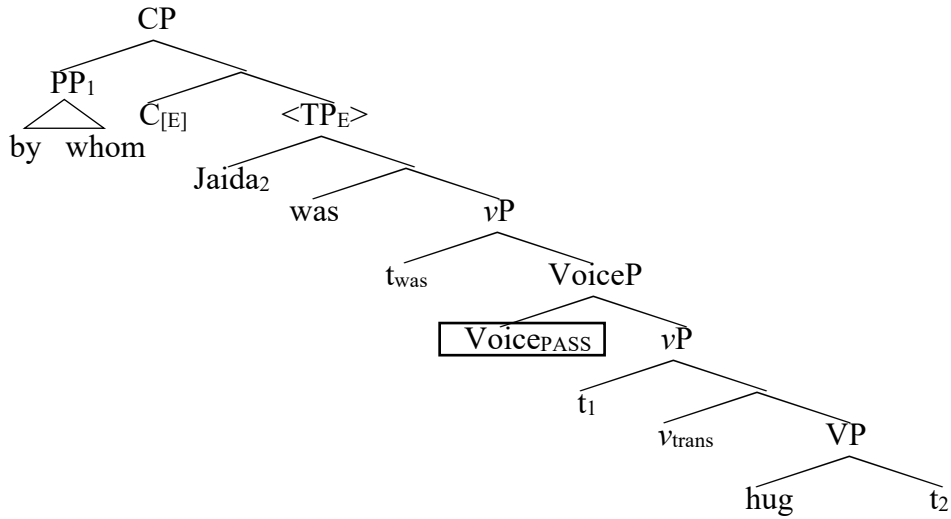
A similar issue leads to the ill-formed status of Passive-Active mismatches, as shown by the derivations below:

(6)   *Voice mismatches in sluicing are banned; *Passive-Active* (based on Merchant 2013: 92)

    a. *Jaida was hugged, but we don't know who.

b. *Antecedent*

```
                TP
        ┌────────┴────────┐
      Jaida₁     ┌─────────┴─────────┐
               was              vP
                        ┌────────┴────────┐
                      t_was          VoiceP
                              ┌──────────┴──────────┐
                        ┌─Voice_PASS─┐        vP
                                          ┌──────┴──────┐
                                        Arg        ┌────┴────┐
                                              v_trans        VP
                                                      ┌──────┴──────┐
                                                    hug           t₁
```

c. *Target clause*

```
                  CP
          ┌────────┴────────┐
       who₁        ┌─────────┴─────────┐
                 C_[E]            <TP_E>
                          ┌──────────┴──────────┐
                        t₁         ┌─────────┴─────────┐
                                  T            VoiceP
                                        ┌────────┴────────┐
                                  ┌─Voice_ACT─┐      vP
                                                  ┌──────┴──────┐
                                                t₁        ┌─────┴─────┐
                                                    v_trans         VP
                                                            ┌──────┴──────┐
                                                          hug          Jaida
```
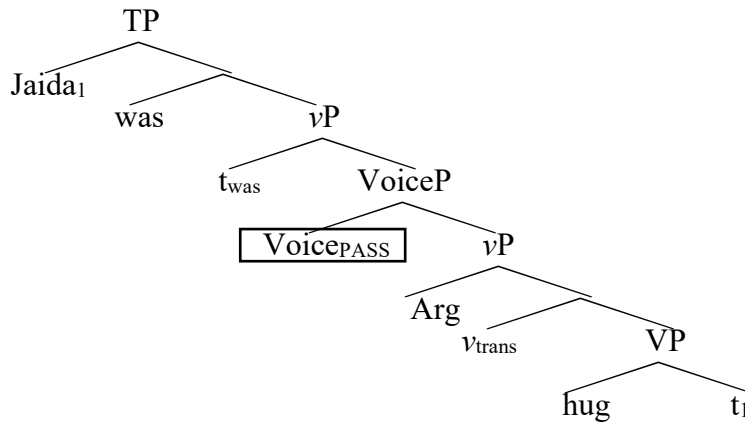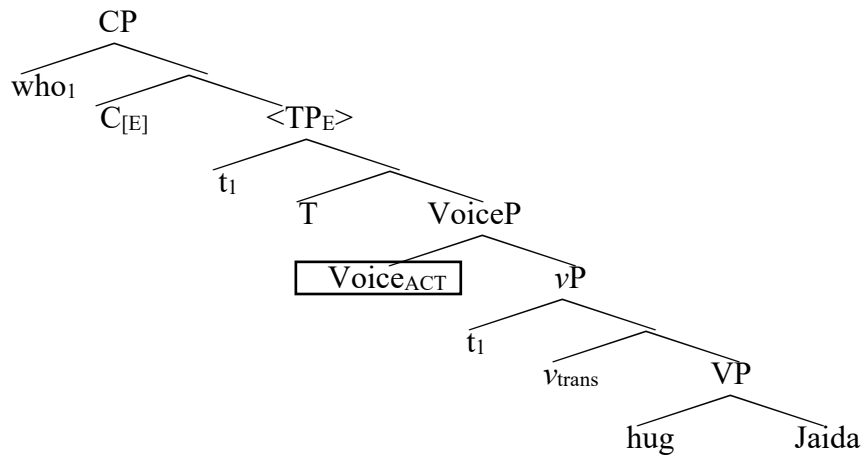
Furthermore, the condition in (4) can also account for why voice mismatches are licit in cases of ellipses targeting a lower chunk of structure, such as VP ellipsis (henceforth VPE). Assuming that the ellipsis site in VPE is lower than VoiceP, then the condition in (4) is satisfied in such cases (Merchant 2013a; see Kehler 2002, Kertz 2013, Kehler 2019 and chapter 1 here for discussion on pragmatic conditions on VP ellipsis remnants that must be met in order for the mismatches to be acceptable; see Sailor 2014 for a syntactico-centric approach to the kind of facts discussed in Kertz 2013):

(7)   *Voice mismatches in VP ellipsis are allowed* (Merchant 2013)

     a. The janitor must remove the trash wherever it is apparent that it should be <removed>.

     b. The system can be used by anyone who wants to <use the system>.

     c. This problem was to have been looked into, but nobody did <look into this problem>.

     d. I have implemented it with a manager, but it doesn't have to be <implemented>.

Let us observe, then, the structural contrast between sluicing and VPE in relation to the mismatches we surveyed. Assuming that VPE is ellipsis of the complement of Voice^0, then the ellipsis site does not contain Voice^0 in VPE. An Active-Passive VPE mismatch is illustrated below, with irrelevant details omitted:

(8)   *Voice mismatches in VPE are allowed*; Active-Passive (adapted from Merchant 2013: 91)

     a. I have implemented it with a manager, but it doesn't have to be.

     b. *Antecedent clause*

c. *Target clause*

TP
- it$_2$
- doesn't
  - have
    - to
    - vP
      - be
      - VoiceP
        - [Voice$_{PASS [E]}$]
        - <vP>
          - Arg
            - v$_{trans}$
            - VP
              - implement
              - t$_2$
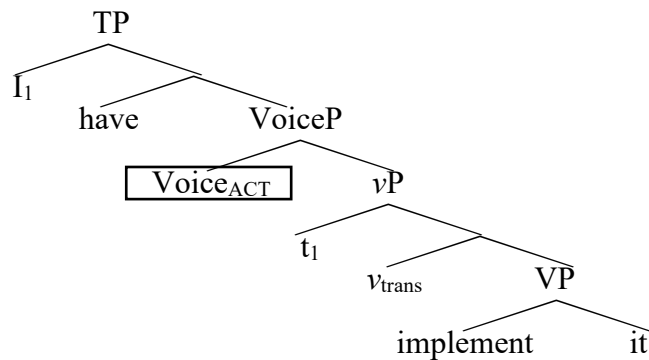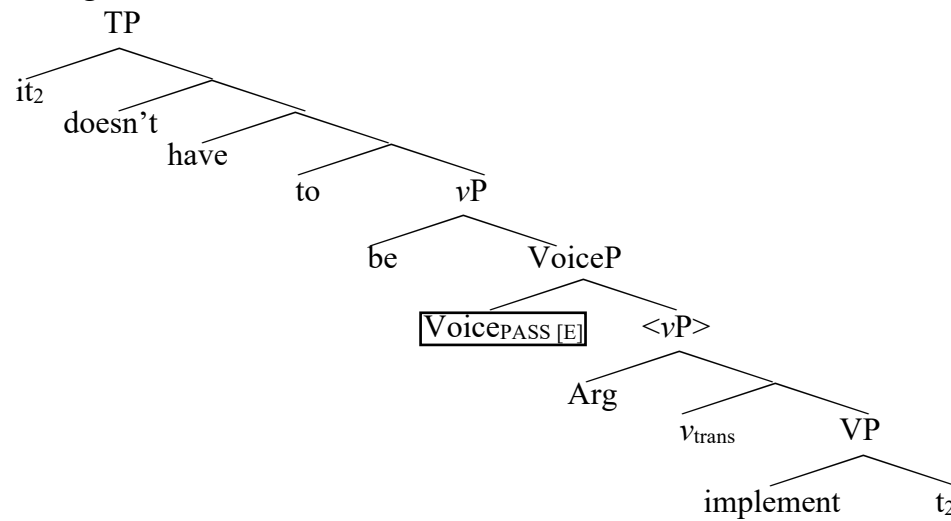
A Passive-Active VPE mismatch is allowed for the same reason—Voice^0 is not included in the ellipsis site:

(9)   *Voice mismatches in VPE are allowed*; Passive-Active (adapted from Merchant 2013: 90)

a. This problem was to have been looked into, but obviously nobody did.
b. *Antecedent clause*

[$_{DP}$ this problem]$_1$ was to have vP
- been
- VoiceP
  - [Voice$_{PASS}$]
  - vP
    - Arg
      - v$_{trans}$
      - VP
        - look into
        - t$_1$
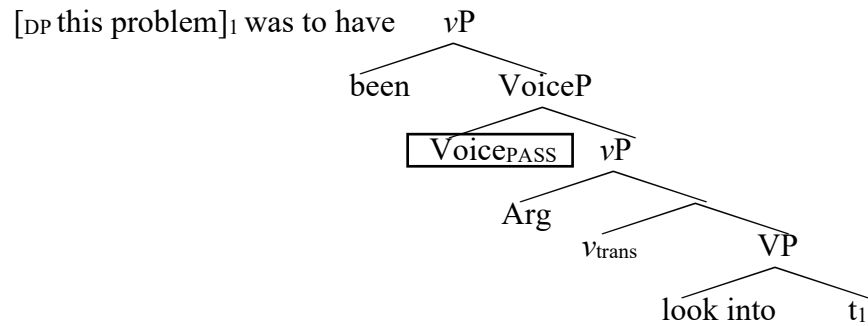
c. *Target clause*
but obviously

```
                    TP
              ┌─────┴─────┐
         nobody₁       ┌──┴──┐
                      did   VoiceP
                         ┌────┴────┐
                   [Voice_ACT[E]]  <vP>
                              ┌─────┴─────┐
                             t₁        ┌──┴──┐
                                    v_trans  VP
                                         ┌────┴────┐
                                    look into    DP
                                            ┌─────┴─────┐
                                          this       problem
```
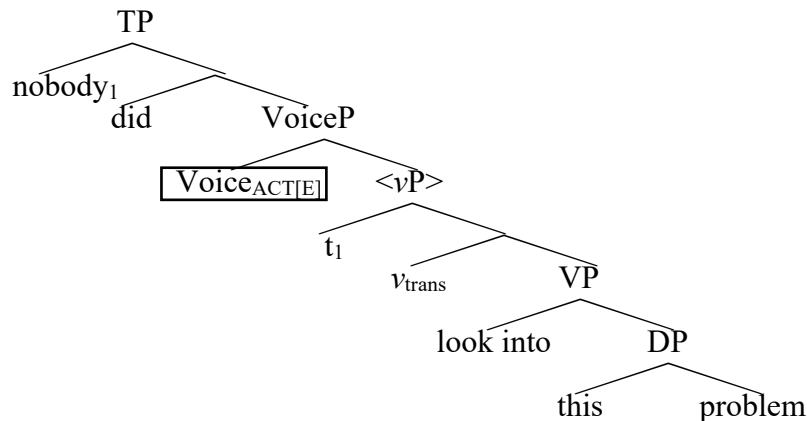
This is how the asymmetrical behavior between sluicing and VPE with regards to voice mismatches is derived in Merchant 2013a.

Consider now the examples below, which show that argument structure alternations are impossible under VPE (Merchant 2013a; (10)b from Johnson 2004; (10)d from Sag 1976; (10)f from Houser et al. 2007; see Levin 1982, Chung et al. 1995, 2011, Chung 2013a for more examples):

(10)   *Argument structure mismatches in VPE*

    a. This can freeze. Please freeze this.
    b. *This can freeze. Please do.
    c. Bill melted the copper vase, and the magnesium vase melted, too.
    d. *Bill melted the copper vase, and the magnesium vase did, too.
    e. Maria still tried to break the vase even though it wouldn't break.
    f. *Maria still tried to break the vase even though it wouldn't.

The data below illustrate that the same flavor of alternation disallowed above under VPE is also banned under sluicing:

(11)   *Argument structure mismatches in sluicing* (Merchant 2013a: 99-100)

    a. They embroidered something with peace signs.
    b. They embroidered peace signs on something.
    c. *They embroidered something with peace signs, but I don't know [what on]₁ <they embroidered peace signs t₁>.

d. *They embroidered something on their jackets, but I don't know [with what]$_1$ <they embroidered their jackets t$_1$>.

(*On image impression reading of 'with what'*)

The way to account for the ill-formed status of argument structure mismatches in (10) and (11) is straightforward: the locus of the featural mismatch is v^0, which is inside the ellipsis site in both types of cases (see Merchant 2013a: 100-101).

Let us summarize the key takeaways from our discussion so far. First, mismatches at the VoiceP/vP level reveal a few empirical generalizations: (i) voice mismatches are disallowed under sluicing but (ii) are allowed in VPE, whereas (iii) argument structure alternations (encoded on v^0) are disallowed across elliptical constructions. This empirical observation suggests strongly that there must be a syntactic component to the calculation of identity under ellipsis. Looking specifically at the voice data, an active clause and its passive counterpart are truth-conditionally equivalent, so a purely semantic formulation of the identity requirement based on mutual entailment (Merchant 2001) overgenerates, ruling the mismatches in, contrary to fact. Second, we also observed that it is not the case that voice mismatches *in general* are disallowed under ellipsis. The contrast between sluicing and VPE is crucial in showing this—if we assume that the VoiceP layer is not included in the silence in VPE, then a strict syntactic identity requirement like (4) derives the asymmetry straightforwardly. Voice can mismatch under VPE, as opposed to sluicing, because the relevant head that would mismatch is not included in the calculation of identity in VPE.

However, the empirical claim that voice mismatches are disallowed in sluicing has been made on the basis of only a handful of languages.[11] Crucially, these languages do not possess a rich inventory of voice alternations whose manipulation under ellipsis could allow for more fine-grained empirical distinctions. In this chapter, I expand the empirical domain that is brought to bear on theories of syntactic identity. To be precise, I will discuss novel sluicing data from Kaqchikel, a Mayan language of the K'ichean branch spoken by around 500,000 people in Guatemala (ISO 639-3).[12] Even though the Mayan languages are the best documented language family in the Americas, ellipsis has never been investigated in detail for any Mayan language.[13] The data presented here are the result of in-person elicitations carried out over a two-year period (2018-2020) with three consultants who are speakers of the Patzún dialect (see Patal Majzul et al. 2000 for dialectal variation in Kaqchikel).

---

[11] The most notable exceptions are Chung 2013a for Chamorro and Potsdam 2003, 2007 for Malagasy. I discuss these languages in chapter 5.

[12] https://www.ethnologue.com/language/cak

[13] The phenomenon is not discussed in Aissen et al. (2017)'s survey of the Mayan languages, nor Coon (2016)'s overview, and it is not discussed in any of the reference grammars and theoretical papers that I have been able to survey. There are a few exceptions, though, that deserve mention. Dayley (1985) briefly discusses a gapping-like phenomenon in San Juan la Laguna Tz'utujil, without committing to any theoretical analysis of the data. Erschler (2018: 183) affirms that there is sluicing in Poqomchi, but provides no data. Anderbois 2008 discusses sluicing in Yukatek and active-antipassive mismatches. Scott Anderbois (p.c.) notes, however, that it is unclear whether the antipassive voice in Yukatek is fully productive, rendering the discussion in the manuscript inconclusive for our current purposes. Finally, Otaki et al. 2013 and Erlewine & Levin 2017 discuss argument ellipsis in Kaqchikel.

A particular class of mismatches will be the focus of our discussion. In Kaqchikel, an active or passive clause can mismatch with a clause in the Agent Focus voice (henceforth AF) under sluicing, and vice-versa (see 2.2.3.5 for some comments on the status of AF as a voice). Representative examples are given below. In (12), the antecedent verb displays active voice morphology, while the sluice shows AF. In (13), the antecedent verb displays AF voice morphology, while the sluice shows active voice. I will provide ample justification in what follows that the proper manipulation of the *wh*-remnant ensures that the voice specification in the sluices must be as specified in the illustrative examples below. Notice also, as a matter of initial, morphological observation, that AF morphology indeed takes the place of active morphology, rather than appearing alongside it.

(12) *Active-AF mismatch*
Jun winäq x-Ø-Ø-tzak-**ij**   la   che'. Man w-etama-n   ta   achike
one person COM-B3S-A3S-cut-**ACT** DEM tree   NEG A1S-know-PRF NEG which
winäq$_1$ <x-Ø-tzak-**in**   la   che' t$_1$>.
person   COM-B3S-cut-**AF** DEM tree
'Some person cut that tree. I don't know which person.'

(13) *AF-Active mismatch*
Ja   ri   xta Maria x-Ø-kam-sa-**n**   jun chiköp… Pero man w-etama-n
FOC DET CLF Maria COM-B3S-die-CAUS-**AF** one animal   but NEG A1S-know-PRF
ta   achike chiköp$_1$ <x-Ø-u-kam-sa-**j**   t$_1$ ri   xta Maria>.
NEG which animal   COM-B3S-A3S-die-CAUS-ACT   DET CLF Maria
'MARIA killed an animal... But I don't know which animal.'

In order to account for data like the above—which stand in contrast to languages like English, which disallow voice mismatches under sluicing—I will put forward an analysis that will have two components. This proposal will form the analytical backbone of the dissertation moving forward.

First, I will show that AF instantiates a clause in which the VoiceP layer has been removed through the course of the derivation via the operation of Exfoliation (Pesetsky 2019). The result for Kaqchikel is the following:

(14)   *AF in Kaqchikel*
       An AF clause is a clause with no VoiceP layer.

Based on the above and the results of our investigation into allowable voice mismatches, I will argue for the following syntactic identity condition on ellipsis:

(15)   *Syntactic identity in ellipsis*
       The antecedent and material properly contained within the ellipsis site must be featurally non-distinct.

Given the analysis sketched in (14), it will be the case that the allowable voice mismatches in Kaqchikel comply with (15). In other words, none of the allowable mismatches instantiate a *clash* between conflicting $Voice^0$ heads. Rather, those data involve mismatches between a clause containing $Voice^0$ and another lacking $Voice^0$ altogether, a configuration allowed by (15). Additionally, we will see that cases that *do* involve a clash between $Voice^0$ heads in Kaqchikel are disallowed, in line with broader cross-linguistic generalizations. Importantly, we will focus on the *syntactic* aspect of the identity condition, abstracting away from the possibility that an additional, semantic and/or pragmatic component is required. Furthermore, the empirical coverage of the condition in (15) will be expanded in chapters 3 and 4, which will respectively discuss mismatches cross-linguistically and in the TP/Mittelfeld and the nominal domain.

2.2 Kaqchikel morphosyntax and sluicing

In this section, I lay out the central empirical facts that will be analyzed in this chapter. First, I provide the basics of Kaqchikel morphosyntax. I then describe sluicing-like

constructions in the language, arguing on the basis of four diagnostics that they should be analyzed as instantiating true sluicing, as opposed to pseudosluicing (defined later as a reduced cleft). I conclude this section by describing the available and unavailable voice mismatches under sluicing in Kaqchikel. Unless otherwise noted, all the data come from my elicitations carried out in 2018-2019 with three native speaker consultants of the Kaqchikel dialect spoken in Patzún, in the department of Chimaltenango.

2.2.1 *Kaqchikel morphosyntax*

Kaqchikel is a head-marking, ergative-absolutive VOS language (England 1991, García Matzar & Rodríguez Guaján 1997, Patal Majzul et al. 2000, Clemens & Coon 2018). Nominals do not show case marking, but verbal agreement reveals the argument alignment: the subject of a transitive clause is indexed with ergative/genitive agreement on the verb, while the object of a transitive clause and the single argument of an intransitive predicate are indexed with absolutive agreement. In the Mayanist literature, the ergative agreement paradigm is called Set A, while the absolutive agreement paradigm is called Set B. I follow this labelling convention throughout. As shown in (16), the subject of a transitive is indexed by Set A agreement on the verb, while the object is indexed with Set B agreement. Example (17) shows that the single argument of an intransitive is indexed with Set B agreement.

(16) *Transitive clause*
X-Ø-u-to'   jun ala'   jun achi.
COM-B3S-A3S-help a   young.man a   man
'A man helped a young man.'   (Patal Majzul et.al. 2000: 141)

(17) *Intransitive clause*
X-Ø-tzaq   ri ti a   Francisco.
COM-B3S-fall DET DIM CLF Francisco
'The boy Francisco fell.'   (Patal Majzul et.al. 2000:115)

Set A agreement also co-indexes possessors, as shown below (see Kenstowicz 2013 for morpho-phonological divergence between Set A co-indexing ergative versus genitive agreement in the 1sg):

(18)   *Set A co-indexes possessors*
  a. nu-way
   A1S-tortilla
   'my tortilla'
  b. a-te'
   A2S-mother
   'your mother'
  c. ru-tz'i'
   A3S-dog
   'his/her dog'

The definiteness, animacy, and information structure status of both arguments in a transitive clause affect the felicity of different word orders, but I abstract away from this in what follows (see García Matzar & Rodríguez Guaján 1997, Patal Majzul et al. 2000, Clemens & Coon 2018). I simply note that while VOS is traditionally analyzed as the unmarked order, topicalization of the subject in pre-verbal position is ubiquitous, yielding SVO (see Can Pixabaj & England 2011 for a discussion of information structure in closely-related K'iche' and Aissen 2017a for Mayan in general).

(19)   *SVO ordering*
  Ri ak'wal x-Ø-u-pax-ij   ri b'ojo'y.
  DET boy COM-B3S-A3S-break-ACt DET pot
  'The boy broke the pot.'   (adapted from Patal Majzul et.al 2000: 143)

The table below provides both Set A and Set B paradigms for the language. There is dialectal variation regarding the exact form of these markers, but I abstract from this in what follows (see Patal Majzul et al. 2000: 47-51 for details; see Kenstowicz 2013 on Set A 1sg):

(20)   *Agreement morphemes in Kaqchikel*

Person/Number	Set A (ERG/GEN)		Set B (ABS)
	_C	_V	
1S	nu-	w- (GEN) nw- (ERG)	in-
2S	a-	aw-	at-
3S	ru-	r-	Ø
1P	qa-	q-	oj-
2P	i-	iw-	ix-
3P	ki-	k-	e-

In Patzún Kaqchikel, 3S set A agreement can be dropped in the presence of a 3S (null) set

B marker (Patal Majzul et.al. 2000: 69). Consider the example below, where we observe a

Ø- morpheme instead of the expected 3S set A *u-*:

(21)   *3S set A can be null in the presence of 3S person*
        Jun winäq   x-Ø-Ø-tzak-ij   la   che'.
        one person COM-A3S-E3S-cut-ACT DEM tree
        'A person cut that tree.'

Moving on, let us discuss a lexical category found in Kaqchikel that is known in the

literature as a relational noun (henceforth RN). RNs serve several functions, similar to

adpositions in other languages. For instance, they encode spatial relations (22)a. In

descriptive terms, they also introduce the oblique, demoted arguments of passives (22)b

and antipassives (22)c. RNs display set A agreement controlled by their argument. In the

examples below, all RNs display set A3S agreement:

(22)   *Relational nouns in Kaqchikel*
        a. ru-xikin jay
           A3S-RN   house
           'bathroom' (*Literally* 'the house's ear')
        b. Ri   tz'i' x-Ø-muq-utäj   **r-uma** ri   achi.
           DET dog COM-B3S-bury-PAS **A3S-RN** DET man
           'The dog was buried by the man.'   (Patal Majzul et.al. 2000:153)

c. Ja   ri   ak'wal x-Ø-chap-o   **r-ichin** ri   tz'i'.
   FOC DET boy   COM-B3S-grab-APO **A3S-RN** DET dog
   'The BOY grabbed the dog.'   (Patal Majzul et.al. 2000:154)

Kaqchikel is a *wh*-movement language: All types of *wh*-phrases surface in the left periphery. *Wh*-interrogation exhibits island effects, as expected if it involves movement. I provide two island-violating examples below, to illustrate (see Mendes & Ranero 2021 for a discussion of the particle that arises with adjunct A'-extraction):

(23)   *Island violation: extraction from an if-clause*
   a. Rat n-Ø-a-tïj   re   ik   re'   si ma Juan n-Ø-u-k'ay-ij
      2S   INC-B3S-A2S-eat DEM chili.pepper DEM if CLF Juan INC-B3S-A2S-sell-ACT
      ri   ru-tz'i'.
      DET A3S-dog
      'You will eat this chili pepper if Juan sells his dog.'

   b. Achike$_1$ n-Ø-a-tïj   t$_1$ si ma Juan n-Ø-u-k'ay-ij   ri ru-tz'i'?
      what   INC-B3S-A2S-eat   if CLF Juan INC-B3S-A2S-sell-ACT DET A3S-dog
      'What will you eat if Juan sells his dog?'

   c. *Achike$_1$ n-Ø-a-tïj   re   ik   re'   si ma Juan
      what   INC-B3S-A2S-eat DEM chili.pepper DEM if CLF Juan
      n-Ø-u-k'ay-ij   t$_1$?
      INC-B3S-A2S-sell-ACT
      Intended: 'What is the thing/creature such that you will eat this chili pepper if
               Juan sold that thing/creature?'

(24)   *Island violation: extraction from a complex NP*
   a. Yïn x-Ø-in-tz'ët   ri achin ri   x-Ø-tj-o   jun aq pa
      1S COM-B3S-A1S-see DET man   REL COM-B3S-eat-AF one pig PREP
      way-b'äl jay.
      eat-INST house
      'I saw the man who ate a pig at the restaurant.'

   b. Ankuchi$_1$ x-Ø-a-tz'ët   (wi) t$_1$ ri   achin ri   x-Ø-tj-o   jun   aq?
      where   COM-B3S-A2S-see FP   DET man   REL COM-B3S-eat-AF one pig
      'Where did you see the man who ate a pig?'

   c. *Ankuchi$_1$ x-Ø-a-tz'ët   ri   achin ri   x-Ø-tj-o   (wi) jun aq t$_1$?
      where   COM-B3S-A2S-see DET man   REL COM-B3S-eat-AF FP   one pig
      Intended: 'Which is the place such that you saw the man who ate a pig at that
                  place?'

I will take for granted from this point forward that *wh*-interrogation involves movement in Kaqchikel (see Erlewine 2016 for additional evidence).

Let us now turn to the different voices that we find in Kaqchikel. There are five distinct voices in the language: (i) active, (ii) passive, (iii) absolutive antipassive (henceforth AP), (iv) oblique antipassive (henceforth AP<small>OBL</small>), and (v) Agent Focus (AF). I discuss each of these in turn but leave a more comprehensive discussion of the properties of AF and AP<small>OBL</small> for sections 2.3 and 2.5 respectively.

An active transitive clause contains two full, non-oblique arguments. The verb displays Set B and A morphology co-indexing the internal and external arguments respectively. There is also a morphological reflex of the active voice in some cases, such as (25)b, where the stem-final *–Vj* appears (we will return to the details of the syntactic structure and morphological make-up of different verbal stems in section 2.3.3):

(25) *Active voice in Kaqchikel*
     a. X-**e-ki**-chäp   ak'wala' ri   ixöqi'.
        COM-**B3P-A3P**-grab children   DET women
        'The women grabbed the children.'
                                    (García Matzar & Rodríguez Guaján 1997: 369)
     b. A   Lu'   x-Ø-u-q'et-**ej**   ri   xta Mari'y.
        CLF Pedro COM-B3S-A3S-hug-**ACT** DET CLF Maria
        'Pedro hugged Maria.'   (McKenna Brown. et al. 2006:177)

In the passive voice, the agent of a two-place verb is expressed by an oblique phrase and the theme appears in surface subject position. Verbs in the passive voice display set B agreement only (controlled by the syntactic subject; i.e., the theme) and a stem-final passive suffix. In a passive clause, the agent can be omitted or introduced obliquely via the relational noun *oma/uma*. There are several types of passives in Kaqchikel, with slightly

different interpretations (García Matzar & Rodríguez Guaján 1997). Below I illustrate passives with the -*Vx* suffix and with the -*Vtäj* suffix:[14]

(26) *Passive voice in Kaqchikel*

a. A   Lu'   x-**Ø-u**-q'et-**ej**   ri   xta   Mari'y.
   CLF Pedro COM-**B3S-A3S**-hug-**ACT** DET CLF Maria
   'Pedro hugged María.'

b. Xta Mari'y x-Ø-q'et-**ëx**   r-oma a   Lu'.
   CLF Maria COM-B3S-hug-**PAS** A3S-RN CLF Pedro
   'María was hugged by Pedro.' (adapted from McKenna Brown. et al. 2006:177)

c. X-**e-ki**-chäp   ak'wala' ri   ixoqi'.
   COM-**B3P-A3P**-grab children DET women
   'The women grabbed the children.'
                          (García Matzar & Rodríguez Guaján 1997: 369)

d. X-**e**-chap-**atäj**   ri   ak'wala' k-uma ri   ixoqi'.
   COM-**B3P**-grab-**PAS** DET children A3P-RN DET women
   'The children were grabbed by the women.'
                          García Matzar & Rodríguez Guaján 1997: 370)

Let us move on to the antipassive voices (for an overview of the antipassive cross-linguistically, see Polinsky 2017). In clauses specified for antipassive voice, the theme of a two place predicate is demoted, meaning it is not expressed as a direct object but is instead omitted, incorporated, or appears as an oblique phrase. Antipassive verbs show only Set B morphology co-indexed with the agent. We will discuss two types of antipassives in Kaqchikel: (i) the absolutive antipassive (AP) and (ii) the oblique antipassive (APOBL).[15]

---

[14] Patal Majzul et al. 2000: 169 reports that Patzún Kaqchikel is one of the dialects of the language that uses a full set of 10 vowels; e.g., there is a tense <e> and a lax <ë>, the distinction being orthographically marked via umlaut. The tense-lax distinction is relevant when discussing the reflex of passivization with certain verbs; see section 2.3.3.

[15] In some of the Kaqchikel literature, APOBL and AF are both described as the "focus antipassive"; Spanish: *antipasivo de enfoque* (see e.g., Patal Majzul 2013).

In the AP, the object is dropped altogether (García Matzar & Rodríguez Guaján 1997, Heaton 2017). The AP is used to describe actions that are done habitually. The form of this antipassive suffix is always *–Vn* in Patzún Kaqchikel (in contrast, see below for the different forms of the AP_OBL and AF suffixes depending on the stem of affixation).

(27)   *AP voice*
      a. X-i-tz'et-on.
        COM-B1S-see-AP
        'I saw.'   (adapted from Patal Majzul 2013: 39)

      b. N-Ø-ki-tïk   ixim.
        INC-B3S-A3S-plant corn
        'They are planting corn.'

      c. Y-e-tik-**on**.
        INC-B3P-plant-**AP**
        'They are planting.'

One important aspect of clauses in the AP voice is that the theme cannot be A'-extracted:

(28)   *AP with IA extraction*
      a. Y-e-tik-on.
        INC-B3P-plant-AP
        'They plant (corn).'

      b. *Achike ixim y-e-tik-**on**?
          what   corn INC-B3P-plant-**AP**
        Intended: 'What corn do they plant?'

We will return to this fact when we discuss the interaction of the AP and active voice under sluicing.

Let us turn now to the AP_OBL. In this voice, the theme can be omitted or introduced in an oblique phrase via the relational noun *ichin*. The form of the AP_OBL suffix varies depending on the morphological make-up of the stem. In Kaqchikel, most bare roots are CVC (Bennett 2016, Coon 2016). With CVC roots, the suffix is *-o* or *–u*. In contrast, if the AP_OBL suffix attaches to a derived stem (i.e., a stem which in the active voice would be composed of a CVC root and additional suffixes like the causative *–(i)sa* or the active

transitive -Vj), it surfaces as -*Vn* (most frequently as –*on* or –*un)*. The latter variant of the suffix is identical to the (non-oblique) AP.

(29)　　*AP*OBL *voice*

    a. X-Ø-u-chüp　　　ri　q'aq' ri　jäb'.
       COM-B3S-A3S-put.out DET fire　DET rain
       'The rain put out the fire.'

    b. Ja　ri　jäb' x-Ø-chup-**u**　　　　r-ichin ri　q'aq'.
       FOC DET rain COM-**B3S**-put.out-**APO** A3S-RN　DET fire
       'The rain put out the fire.'　　　　(García Matzar & Rodríguez Guaján 1997: 374)

Following observations made in Heaton 2017, we will return to some properties of the APOBL in section 2.5.1. We will see that these properties make the APOBL diverge from the AP in Kaqchikel and antipassives cross-linguistically.

The final voice in Kaqchikel is Agent Focus (henceforth AF). The literature on AF, a term whose use may vary slightly from language to language across the Mayan family, is extensive (Larsen & Norman 1978, Smith-Stark 1978, Craig 1979, Davies & Sam-Colop 1990, Dayley 1981, Aissen 1999, 2011, among others; see Coon et al. 2014, Aissen 2017b, Coon et al. 2021 for discussion). In Kaqchikel, AF is used exclusively when the subject of a transitive clause is A'-extracted. Since A'-movement of the thematic agent of a transitive clause is typically blocked in Kaqchikel active transitives, Kaqchikel can be categorized as exhibiting syntactic ergativity (Coon et al. 2014, Polinsky 2016a, Erlewine 2016, Douglas et al. 2017, Henderson & Coon 2018). In a manner that parallels the form of the APOBL morpheme, the form of the AF suffix varies depending on the morphological make-up of the verb. With CVC roots, the suffix is -*o* or -*u*, whereas with derived stems, the suffix surfaces as -*Vn* (most frequently as -*on* or –*un)*. In a nutshell, the form of the AF suffix is identical to the APOBL. There are two ways, however, to distinguish between the two voices: (i) the presence of a relational noun introducing the theme in the APOBL, but not

AF, and (ii) the agreement pattern in each construction. In the APOBL, Set B morphology invariably co-indexes the external argument, whereas Set B agreement in AF co-indexes whichever argument is higher in an accessibility person/number hierarchy (see (34)c below). The following examples show that *wh*-extraction of the EA and focus of the EA cannot occur in the active voice. Instead, AF must be used:[16]

(30)   *Ungrammaticality of EA wh-extraction with active voice*
        a. *Achike x-Ø-u-tïj   nu-way?
            who   COM-B3S-A3S-eat A1S-tortilla
            Intended: 'Who ate my tortillas?'
        b. Achike x-Ø-tj-o   nu-way?
            who   COM-B3S-eat-AF A1S-tortilla
            'Who ate my tortillas?'

(31)   *Ungrammaticality of EA focus with active voice[17]*
        Q: 'Who bought the flowers?'
        a. Xa   xe   ma Juan x-Ø-loq'-o   ri   kotz'i'j.
            EMPH only CLF Juan COM-B3S-buy-AF DET flower
            'Only Juan bought the flowers.'
        b. *X-Ø-u-löq'   ri   kotz'i'j xa   xe   ma Juan.
            COM-B3S-A3S-buy DET flower EMPH only CLF Juan
            Intended: 'Only Juan bought the flowers.'

If the EA of a transitive clause is an existential indefinite, AF is also required (see Erlewine 2016 for evidence that this configuration involves movement):

---

[16] Not all Mayan languages are syntactically ergative, and certain dialects of Kaqchikel have been reported to lack the extraction restriction in a subset of A'-environments (Heaton et. al. 2016, Douglas et.al. 2017). Crucially for our present purposes, all the speakers consulted for this work display a uniform extraction restriction.

[17] The *xa* particle is described as "devaluative" or "emphatic" (*Spanish*: desvalorativa, enfática) in Patal Majzul 2013: 47, while the *xe* particle is described as "only" or "inmediative" (*Spanish*: solo, inmediativo). The entry for *xa xe* on p. 533 translates the term as 'only' (*Spanish*: solamente).

(32)   *Existential indefinite and AF*

    a. K'o   jun x-Ø-loq'-o   ri kotz'i'j.
       EXIST one COM-B3S-buy-AF DET flower
       'Someone bought flowers.'

    b. *K'o   jun x-Ø-u-löq'   ri   kotz'i'j.
       EXIST one COM-B3S-A3S-buy DET flower
       Intended: 'Someone bought flowers.'

A related observation, which will figure prominently in the analysis proposed in section 2.3, is that A'-movement of any element other than the EA of a transitive is incompatible with AF. In other words, if we A'-extract internal arguments or adjuncts, it is impossible to use AF. This is demonstrated below:

(33)   *AF incompatible with object/adjunct A'-movement*

    a. Achike x-Ø-u-tïj   ma Juan?
       what   COM-B3S-A3S-eat CLF Juan
       'What did Juan eat?'

    b. *Achike x-Ø-tj-o   ma Juan?
       what   COM-B3S-eat-AF CLF Juan
       Intended: 'What did Juan eat?'

    c. K'o   jun x-Ø-loq'-o   ri   aq. Aw-etama-n   achike r-uma
       EXIST one COM-B3S-buy-AF DET pig A2S-know-PERF WH   A3S-RN
       x-Ø-u-löq'?
       COM-B3S-A3S-buy
       'Someone bought the pig. Do you know why they bought it?'

    d.*K'o jun x-Ø-loq'-o   ri   aq. Aw-etama-n   achike r-uma
       EXIST one COM- B3S-buy-AF DET pig A2S-know-PERF WH   A3S-RN
       x-Ø-loq'-o?
       COM-B3S-buy-AF
       Intended: 'Someone bought the pig. Do you know why they bought it?'

    e. Ankuchi x-Ø-u-löq'   (wi) ri   kotz'i'j?
       where   COM-B3S-A3S-buy FP   DET flower
       'Where did s/he buy the flowers?'

    f. *Ankuchi x-Ø-loq'-o   (wi) ri   kotz'i'j?
       where   COM-B3S-buy-AF FP   DET flower
       Intended: 'Where did s/he buy the flowers?'

AF also displays a cluster of properties that distinguish it from the other voices in Kaqchikel. An overview of these properties is given below:

(34)   *Properties of AF clauses*
        a. ERG agreement does not surface.
        b. Only ABS agreement surfaces.
        c. ABS agreement is controlled either by the EA or the IA, depending on an accessibility hierarchy 1, 2 > 3PL > 3S
        d. AF cannot be used if both arguments are local persons (i.e., 1st and 2nd person).

The details of these properties of AF will be discussed in section 2.4. With this background in mind, we are now ready to delve into sluicing in Kaqchikel.

2.2.2 Sluicing in Kaqchikel

Sluicing is a type of clausal ellipsis that was first discussed in the seminal work by Ross 1969a. Sluicing involves ellipsis of an entire clause, except a *wh*-remnant.

(35)   *Sluicing*
        Someone in the department likes metal music, but no one knows who$_1$ <t$_1$ likes metal music>.

Consider a Kaqchikel example which looks on the surface like sluicing. For now, I will refer to this type of data as a sluicing-like construction (SLC):

(36)   *Kaqchikel SLC; EA wh-phrase remnant*
        K'o jun x-Ø-loq'-o=pe   k'iy knaq'… man w-etama-n
        EXIST one COM-B3S-buy-AF=DIR many bean   NEG A1S-know-PRF
        ta   achike <x-Ø-loq'-o=pe   k'iy knaq'>.
        NEG who   COM-B3S-buy-AF many bean
        'Someone bought a lot of beans… I don't know who <bought a lot of beans>.'

Let us lay out some terminology that will be used in the description and discussion of the data. These terms are not intended, at this stage, as formal components of any particular analysis, but rather as a means with which to talk about various elements of a sluicing example, and other elliptical constructions as well:

(37)   *Terminology*
    a. **Antecedent:** the first clause in (36); '*Someone bought a lot of beans*.'
    b. **Target clause**: the second clause in (36); *'I don't know who.'*
    c. **Ellipsis site/sluice**: the elided part in (36); *<bought a lot of beans>*.
    d. **Remnant:** material outside the ellipsis site; i.e. '*who*' in (36) is the *wh*-remnant
    e. **Correlate:** material in the first clause which corresponds to some element of the remnant; i.e., '*someone*' is the correlate of the *wh*-remnant '*who*' in (36).

Moving on, observe that arguments or adjuncts can be *wh*-remnants in Kaqchikel SLCs. In (36), we observed an EA as a *wh*-remnant. The examples given below show an IA *wh*-remnant (38) and a locative *wh*-remnant (39):

(38)   *Kaqchikel SLC; IA wh-phrase remnant*

Ri a   Juan k'o   x-Ø-u-löq'=pe.   Man w-etama-n   ta   achike.
DET CLF Juan EXIST COM-B3S-A3S-buy DIR NEG A1S-know-PRF NEG what
'Juan bought something. I don't know what.'

(39)   *Kaqchikel SLC; locative wh-phrase remnant*

A: Xa   xe   ri ma Pedro x-Ø-loq'-o   ri   kotz'i'j.
   EMPH only DET CLF Pedro COM-B3S-buy-AF DET flower
   'Only Pedro bought the flowers.'

B: Aw-etama-n   ankuchi?
   A2S-know-PRF   where
   'Do you know where?'

Our first task is to determine whether these examples instantiate true sluicing, or if they instantiate instead a reduced cleft, which I will refer to as a *pseudosluice* (Merchant 1998).[18] We must determine this in order to ensure that the phenomenon we will focus on involves voice mismatches under *sluicing*, rather than some other phenomenon. Note,

---

[18] Specifically, Merchant 1998 used the term pseudosluice as an analysis of structures that look like sluicing in Japanese. Merchant proposed that the Japanese data do not involve clausal ellipsis; rather, the surface string is a combination of pro-drop and a null copula taking a *wh*-phrase as a predicate (see van Craenenbroeck & Merchant 2013: 745 for some discussion).

however, that even if we determined that the phenomenon involved voice mismatches under *pseudosluicing*, that would not mean that there is no problem to be explained. For instance, Adams & Tomioka (2012) defend a pseudosluicing analysis of SLCs in Mandarin Chinese, but note that voice mismatches are impossible. In other words, we will lay out several arguments that the SLCs discussed here do not instantiate pseudosluicing in order to have a clear idea of the structure underpinning the ellipsis site, *not* because a pseudosluicing analysis would be a possible alternative analysis that would render our analytical conclusions moot.

Let us observe the two different structures that underlie the competing analyses we will entertain for SLCs. I assume that sluicing involves *wh*-extraction followed by TP-ellipsis (Ross 1969a, Merchant 2001; in Kaqchikel, this amounts to InflP ellipsis—see (92) below for the basic clause structure I assume). The structure from which the *wh*-phrase is extracted is elided, leaving the correlate on the surface. A sketch of this derivation is seen in (40):

(40)   *Sluicing structure*
        I don't know who <_ ~~bought many beans~~>.

A reduced cleft/pseudosluice, in contrast, involves extraction of the *wh*-pivot of a cleft. The clefted structure would not surface in a Kaqchikel pseudosluice:

(41)   *Pseudosluicing structure*
        I don't know who it was <_ ~~that bought many beans~~>.

In a nutshell, we need to determine whether the elided structure in Kaqchikel SLCs involves a sluice, as in (40), or a pseudosluice as in (41).

The problem is that it is not so easy to distinguish sluicing from a reduced cleft in Kaqchikel. The reason is twofold: (i) Kaqchikel is a null argument language with no (overt)

expletives and (ii) Kaqchikel has no overt copula.[19] Therefore, the Kaqchikel counterparts of (40) and (41) would be string-identical on the surface. Below, I sketch what the structure would look like for either analysis.

(42)　*Sluicing analysis*

K'o　jun　x-Ø-loq'-o=pe　　　　k'iy　knaq'… man w-etama-n　　ta　achike₁
EXIST one COM-B3S-buy-AF =DIR many bean　　NEG A1S-know-PRF NEG who
< [InflP t₁ x-Ø-loq'-o=pe　　　　k'iy　knaq']>.
　　　　COM-B3S-buy-AF=DIR many bean
'Someone bought a lot of beans… I don't know who <bought a lot of beans>.'

(43)　*Pseudosluice (cleft) analysis*

K'o　jun　x-Ø-loq'-o=pe　　　　k'iy　knaq'… man w-etama-n　　ta　achike
EXIST one COM-B3S-buy-AF=DIR many bean　　NEG A1S-know-PRF NEG who
pro　Ø　<_ ri　x-Ø-loq'-o=pe　　　　k'iy　knaq'>.
EXPL COP　　REL COM-B3S-buy-AF=DIR many bean
'Someone bought a lot of beans… I don't know who it was <that bought a lot of beans>.'

In the next subsections, I will assess the status of SLCs using four diagnostics:

(44)　*Diagnostics to distinguish sluicing from pseudosluicing*
　　a. Compatibility with aggressively non-D-linked *wh*-phrases
　　b. Availability of adjunct sprouting
　　c. Availability of *mention some* modification
　　d. Availability of *else* modification

I will argue that the result of all four diagnostics favor a sluicing (rather than pseudosluicing) analysis of Kaqchikel SLCs. Crucially, I will also show that there is an *independent* construction in the language that patterns with a reduced cleft/pseudosluicing, showing the opposite behavior with respect to the same four diagnostics. Thus, we will conclude that the construction at the center of our discussion truly involves sluicing.

---

[19] Note that Kaqchikel has grammaticalized aspect, not tense. To the best of my knowledge, there is no copula in any aspectual context.

2.2.2.1 Diagnostic 1: Compatibility with aggressively non-D-linked *wh*-phrases

It has been observed that aggressively non-D-linked *wh*-phrases cannot be remnants in sluicing. In contrast, they are felicitous as the pivots of clefts (Pesetsky 1987, Merchant 1998, Merchant 2001):

(45)   *Aggressively non-D-linked wh-phrases cannot be remnants in sluicing*
     a. Someone ate my tortillas, but I don't know who.
     b. *Someone ate my tortillas, but I don't know *who the hell*!

(46)   *Aggressively non-D-linked wh-phrases can be pivots of clefts*
     a. Someone ate my tortillas, but I don't know who it was.
     b. Someone ate my tortillas, but I don't know *who the hell* it was!

Our prediction for Kaqchikel is as follows: If SLCs instantiate true sluicing, then aggressively non-D-linked *wh*-phrases should be unavailable as remnants. Conversely, if SLCs involve a pseudosluice, then an aggressively non-D-linked *wh*-phrase should be possible as a remnant. The data given below show that the Kaqchikel examples pattern with sluicing: An aggressively non-D-linked *wh*-phrase cannot be a correlate, as shown in (47)c with an embedded sluice with a subject *wh*-remnant. Note, crucially, that it is possible in general to use an aggressively non-D-linked *wh*-phrase, as the non-ellipsis example in (47)d shows.

(47)   *SLCs and aggressively non-D-linked wh-remnants (subject wh-remnant)*
     a. K'o   jun x-Ø-tj-o   nu-way,   po man w-etama-n   ta   achike.
        EXIST one COM-B3S-eat-AF A1S-tortilla but NEG A1S-know-PRF NEG who
        'Someone ate my tortillas, but I don't know who.'
     b. K'o   jun x-Ø-tj-o   nu-way,   po man w-etama-n   ta   achike
        EXIST one COM-B3S-eat-AF A1S-tortilla but NEG A1S-know-PRF NEG who
        x-Ø-b'an-o
        COM-B3S-do-AF
        'Someone ate my tortillas, but I don't know who did it.'
     c. *K'o jun x-Ø-tj-o   nu-wäy,   po man w-etama-n   ta   achike kïs.
        EXIST one COM-B3S-eat-AF A1S-tortilla but NEG A1S-know-PRF NEG who   shit
        Intended: 'Someone ate my tortillas, but I don't know who *the shit*.'

d. K'o   jun x-Ø-tj-o   nu-wäy,   po man w-etama-n   ta   achike kïs
   EXIST one COM-B3S-eat-AF A1S-tortilla but NEG A1S-know-PRF NEG who   shit
   x-Ø-b'an-o.
   COM-B3S-do-AF
   'Someone ate my tortillas, but I don't know who *the shit* did it.'

The example given below illustrates the same point with a matrix sluice—aggressively

non-D-linked phrases cannot be *wh*-remnants in Kaqchikel SLCs:

(48)   *SLCs and aggressively non-D-linked wh-remnants (subject wh-remnant)*

   a. K'o   jun x-Ø-tj-o   nu-way.   Achike chqa' salad x-Ø-tj-o?!
     EXIST one COM-A3S-eat-AF A1S-tortilla who   also   salty COM-B3S-eat-AF
     'Someone ate my tortilla. *Who the salty ate it*?!

   b.*K'o   jun x-Ø-tj-o   nu-way.   Achike chqa' salad?!
     EXIST one COM-B3S-eat-AF A1S-tortilla who   also   salty
     Intended: 'Someone ate my tortilla. *Who the salty*?!'

Finally, the example given below demonstrates the same point with an adjunct *wh*-

remnant:[20]

(49)   *SLCs and aggressively non-D-linked wh-remnants (adjunct wh-remnant)*

   a. Achoj k'in x-a-xajo'   rat?
     WH   RN COM-B2S-dance 2S
     'Who did you dance with?'

   b. Achoj k'in kïs x-a-xajo'   rat?!
     WH   RN shit COM-B2S-dance 2S
     'Who the *shit* did you dance with?'

   c.   A: Chi   jun aq'a' x-Ø-ajo'   a   Juan.
       PREP one night COM-B3S-dance CLF Juan
       'Juan danced all night.'

---

[20] Despite the fact that the Kaqchikel ordering of *wh*+complement in (49) looks like *swiping* (*Sluicing With In-Situ Preposition In Northern Germanic*; Rosen 1976, Merchant 2002, Sprouse 2005), this is not the same phenomenon. We will observe later that *wh*+complement adjunct questions in Kaqchikel obligatorily invert, in both ellipsis and non-ellipsis contexts (see Broadwell 2006, Ewing 2020). This is not the case in English, where the inversion occurs only in ellipsis contexts.

B: Kan qitzij? Achoj k'in?
INT truth WH   RN
'Really? With who?'

B′: *Kan qitzij? Achoj k'in kïs?
INT truth WH   RN shit
Intended: 'Really?! Who with *the shit*?'

We see therefore that this first diagnostic favors an analysis of Kaqchikel SLCs as true sluicing, as opposed to pseudosluicing.

<u>2.2.2.2 Diagnostic 2: Availability of adjunct sprouting</u>

The second diagnostic involves the availability of adjunct sprouting in SLCs. Sprouting has come to refer to a sluicing configuration in which the *wh*-remnant has no correlate in the antecedent (see Chung et al. 1995, Chung 2006):

(50)   *Sprouting*
      a. Robin fixed the car, but no one knows how <Robin fixed the car>.
      b. Leslie bought some apples. Do you know where <Leslie bought some apples>?
      c. Sam didn't come to school today. Does anyone know why <Sam didn't come to
        school today>?

Some sprouted adjuncts can occur as the sprouted pivots of clefts in English, while others cannot (Merchant 2001, Potsdam 2007). Consider the contrast between sprouting a comitative as the pivot of a cleft (51)B, versus sprouting a locative as the pivot of a cleft (51)B′:

(51)   *Limited sprouting in clefts*
      A: Juan was dancing all night.
      B: ?Oh. Do you know who it was with?
      B′: *Oh. Do you know where it was?

More examples of unacceptable sprouting in this configuration are shown below:

(52)   *Limited sprouting in clefts*
      a. He fixed the car, but I don't know how (*it was).
      b. He fixed the car, but I don't know why (*it was).

c. He fixed the car, but I don't know when (*it was).
d. He's hidden the jewels, but I don't know where (*it is).   (Potsdam 2007: 606)

I have no explanation for why some adjuncts should be impossible and others more acceptable with sprouting in general. Crucially for our purposes, though, the prediction for Kaqchikel is clear. If SLCs involve true sluicing, then adjunct sprouting should be licit across-the-board. In contrast, if SLCs involve pseudosluicing, then adjunct sprouting should be unavailable at least in a subset of cases. As shown by the data below, sprouting *is* available with all kinds of adjuncts. Furthermore, we will see that the behavior of these same adjunct *wh*-remnants is different in the other construction, which patterns with pseudo-sluicing (as opposed to sluicing) in other respects as well (see 2.2.2.5):[21]

(53)   *Manner adverbial sprout in Kaqchikel SLCs*

Iwir,   x-Ø-ajo'   a   Juan. Man w-etama-n   ta   achike modo.
yesterday, COM-B3S-eat CLF Juan.   NEG A1S-know-PRF NEG what   way
'Yesterday, Juan danced. I don't know in what manner.'

(54)   *Reason adverbial sprout in Kaqchikel SLCs*

Xta Andrea x-Ø-u-löq'   k'iy   knaq'. A   Pedro x-Ø-u-kutu-j
CLF Andrea COM-B3S-A3S-buy many bean   CLF Pedro COM-B3S-A3S-ask-ACT
achike ru-ma.
what   A3S-RN
'Andrea bought a lot of beans. Pedro asked why.'

---

[21] It is possible that relational nouns have been reanalyzed as non-agreeing prepositions in the Kaqchikel of Patzún, since speakers do not consistently use set A on RNs (see Patal Majzul et.al. 2000: 48-9; compare (57) with (58)). Additionally, set A morphology does not agree in number for my consultants when the argument of the relational noun is plural 3rd person (in contrast to examples in García Matzar & Rodríguez Guaján 1997). Note that some languages have bona-fide prepositions showing agreement (e.g., Welsh; King 2003, Hirata 2012). I speculate that Patzún speakers have reanalyzed relational nouns as non-agreeing prepositions.

(55) *Temporal adverbial sprout in Kaqchikel SLCs*

A   Juan x-Ø-ajo',   pero man w-etama-n   ta   jampe'.
CLF Juan COM-B3S-dance but   NEG A1S-know-PRF NEG when
'Juan danced, but I don't know when.'

(56) *Locative adverbial sprout in Kaqchikel SLCs*

Xta Andrea   x-Ø-u-löq'   k'iy   knaq'. A   Pedro x-Ø-u-kutu-j
CLF Andrea   COM-B3S-A3S-buy many bean   CLF Pedro COM-B3S-A3S-ask-ACT
ankuchi.
where
'Andrea bought a lot of beans. Pedro asked where.'

(57) *Comitative sprout in Kaqchikel SLCs*

A   Juan x-Ø-Ø-ajo'   jun son, pero man w-etama-n   ta   achoj r-k'in.
CLF Juan COM-B3S-dance a   son but   NEG A1S-know-PRF NEG WH   A3S-RN
'Juan danced a *son*, but I don't know who with.'[22]

(58) *Benefactive sprout in Kaqchikel SLCs*

Xta Andrea x-Ø-u-këm   jun po't,   po man w-etama-n   ta   achoj chin.
CLF Andrea COM-B3S-A3S-knit one güipil but NEG A1S-know-PRF NEG WH   RN
'Andrea knit a *güipil*, but I don't know who for.'[23]

This second diagnostic also favors an analysis of Kaqchikel SLCs in terms of true sluicing,

as opposed to pseudosluicing.

2.2.2.3 "Mention some" modification

A third diagnostic involves *mention some* modification, which refers to the possibility of

modifying the *wh*-remnant with 'for example' or similar modifiers. Merchant (2001) shows

that *wh*-pivots of clefts are incompatible with such modification, a result of the exhaustivity

associated with the cleft.

(59) *Mention-some modification* (Merchant 2001: 122)
A: You should talk to somebody in the legal department for help with that.

---

[22] A *son* is a traditional dance.

[23] A *güipil* is the traditional blouse knit and worn by many Mayan women.

B: Could you tell me who (*it is), for example?
B′: Who (*is it), for example?

The prediction is clear for Kaqchikel: If SLCs involve sluicing, then we should be able to modify the *wh*-remnant in a manner parallel to the English above. The prediction is borne out. The SLC examples given below are translated by consultants into Spanish in a manner akin to the English examples.[24]

(60)   *Mention some modification in Kaqchikel SLCs*

    a.   A: Ta-loq'-o   jun spanïk chin a-te'!
                IMP-buy-SS a   gift   RN A2S-mother
                'Buy a gift for your mother!'
            B: Achike ta   n-Ø-in-löq'?
                what   IRR INC-B3S-A1S-buy
                'What do I buy, for example?'

    b.   A: Ta-loq'-o   jun spanïk chin a-te'!
                IMP-buy-SS a   gift   RN A2S-mother
                'Buy a gift for your mother!'
            B: Achike ta   k'a?
                what   IRR PART
                'What, for example?' *Spanish*: '¿Qué, por ejemplo?'

    c.   A: Ta-loq'-o   jun spanïk chin a-te'!
                IMP-buy-SS a   gift   RN   A2S-mother
                'Buy a gift for your mother!'
            B: Achike ta   jun?
                what   IRR one
                'What, for example?' *Spanish*: '¿Qué, por ejemplo?'

Again, this third diagnostic favors an analysis of SLCs as true sluicing, instead of pseudosluicing.

---

[24] I leave for future work the precise analysis of the particles *ta* and *k'a* as used in (60) (see García Matzar & Rodríguez Guaján 1997 for discussion).

2.2.2.4 "Else-modification"

A fourth diagnostic is similar in spirit to the previous one. Merchant (2001) notes that the modifier *else* is incompatible with the *wh*-pivot of a cleft:[25]

(61)   *Else-modification* (Merchant 2001: 122)
         Harry was there, but I don't know who else (*it was).

The prediction is as follows: If Kaqchikel SLCs instantiate true sluicing, then it should be possible to modify the *wh*-remnant with 'else'. This prediction is correct:

(62)   *Else modification in Kaqchikel SLCs*
         A: Ma Juan x-Ø-u-löq'   äk'.
            CLF Juan COM-B3S-A3S-buy chicken
            'Juan bought chicken.'
         B: Aw-etama-n   achike chmas?
            A2S-know-PRF what   else
            'Do you know what else?'

We see then that this fourth diagnostic, like the previous three, favors an analysis of Kaqchikel SLCs in terms of true sluicing, as opposed to a pseudosluice.

---

[25] Omer Preminger asks me if it is relevant that an expletive 'there' instead of 'it' makes examples akin to (61) acceptable:

    (i) Harry was in the room, but I don't know who else **there** was.

Whatever the reason for the acceptability of (i), the example is not a reduced cleft. In other words, (i) is not a reduced version of (ii), which is degraded in comparison to (iii) (a cleft):

    (ii) ??Harry was in the room, but I don't know who else **there was that was in the room**.

    (iii) Harry was in the room, but I don't know who else **it was that was in the room**.

Since the point here is to distinguish between true ellipsis versus reduced clefts, the proper analysis of examples like (i) (and why they are possible) falls outside the scope of our discussion.

<u>2.2.2.5 An independent reduced cleft/pseudosluice</u>

Alongside the results of the four diagnostics discussed previously, the conclusion that Kaqchikel SLCs involve true sluicing is further supported by the existence of an *independent* construction in the language that patterns with pseudosluicing. This construction involves a *wh*-remnant plus a determiner, being thus similar to *spading* in Dutch (*Sluicing Plus A Determiner In Non-insular Germanic*; van Craenenbroeck 2012).

An example illustrating this construction in Kaqchikel is given below in (63)c. Moving forward, I will refer to this data type as spading for ease of exposition. However, I make no commitment that the analysis of these data should be identical to the analysis of spading more broadly. I provide a cleft-like translation for the spading examples for three reasons: (i) to distinguish SLCs from spading, (ii) because my consultants provide a cleft-like translation for this type of data, and (iii), the interpretive difference between simple *wh*-questions and those with a determiner suggests that a cleft is involved (see (64)).

(63) *Spading in Kaqchikel*

    a.   A: Xa   xe   ri   ma Juan x-Ø-loq'-o   kotz'i'j.
               EMPH only DET CLF Juan COM-B3S-A3S-buy-AF flower
               'Only Juan bought flowers.'
            B: Kan qitzij? Achike kotz'i'j x-Ø-u-löq'?
               INT truth   which flower COM-B3S-A3S-buy
               'Really? Which flowers did he buy?'

    b.   A: Xa   xe   ri   ma Juan x-Ø-loq'-o   kotz'i'j.
               EMPH only DET CLF Juan COM-B3S-A3S-buy-AF flower
               'Only Juan bought flowers.'
            B: Kan qitzij? Achike kotz'i'j?
               INT truth   which flower
               'Really? Which flowers?'

    c.   A: Xa   xe   ri   ma Juan x-Ø-loq'-o   kotz'i'j.
               EMPH only DET CLF Juan COM-B3S-A3S-buy-AF flower
               'Only Juan bought flowers.'

B: Kan qitzij? Achike kotz'i'j ri'?
INT truth   what   flower   DET
'Really? What flowers were they?'

It is relevant to note that there exists an interpretive difference between simple *wh*-questions and those involving a *wh*-phrase plus a determiner. Consider the contrast below. In the question in (64)b, consultants report that the speaker must have a particular person in mind, whose precise identity she has now forgotten. No such implication exists in the simple *wh*-question (64)a.

(64)   *Simple wh-question vs. wh+determiner*
         a. Achike winäq x-Ø-loq'-o   jun monton kotz'i'j?
            which   person COM-B3S-buy-AF one bunch   flower
            'Which person bought a bunch of flowers?'
         b. Achike winäq ri' x-Ø-loq'-o   jun monton kotz'i'j?
            which   person DET COM-B3S-buy-AF one bunch   flower
            'Which person is it that bought a bunch of flowers?'

It is necessary to investigate further what the precise interpretation of these *wh*+determiner questions is. For our purposes, however, it suffices to note that there are two independent elliptical constructions in Kaqchikel: SLCs and spading. Crucially, spading patterns with pseudosluicing, and in opposition to SLCs, with respect to the four diagnostics discussed previously.

     First, aggressively non-D-linked *wh*-phrases *can* be remnants in spading. This contrasts with the SLC data we discussed earlier:

(65)   *Spading in Kaqchikel and aggressively non-D-linked wh-remnant*
         a.   A: Chi jun aq'a' x-Ø-ajo'   a   Juan.
                   PREP one night COM-B3S-dance CLF Juan
                   'Juan danced all night.'
                B: Kan qitzij? Achoj k'in ri'?
                   INT truth   WH   RN   DET
                   'Really? Who was it with?'

b. A: Chi jun aq'a' x-Ø-ajo'   a Juan.
     PREP one night COM-B3S-dance CLF Juan
     'Juan danced all night.'
    B: Kan qitzij? Achoj k'in kïs ri'?
     INT truth WH RN shit DET
     'Really? Who *the shit* was it with?'

c. *K'o jun x-Ø-tj-o   nu-way**.** Achike chqa' salad?!
 EXIST one COM-B3S-eat-AF A1S-tortilla who also salty
 Intended: 'Someone ate my tortilla. Who *the salty*?!

d. K'o jun x-Ø-tj-o   nu-way**.** Achike chqa' salad ri'?!
 EXIST one COM-B3S-eat-AF A1S-tortilla who also salty DET
 'Someone ate my tortilla. Who *the salty* was it?!

Second, spading is incompatible with the sprouting of certain adjuncts, as shown below with a locative adjunct (see section 2.2.2). Compare the unacceptable spading example in (66)b with its SLC counterpart, in (66)a, which involves the very same adjunct (*ankuchi* 'where') and is well-formed.

(66) *Spading in Kaqchikel and sprouting*
  a. Xta Andrea x-Ø-u-löq'   jun tz'i', po man w-etama-n ta ankuchi.
   CLF Andrea COM-B3S-A3S-buy one dog but NEG A1S-know-PRF NEG where
   'Andrea bought a dog, but I don't know where.'
  b. *Xta Andrea x-Ø-u-löq'   jun tz'i', po man w-etama-n ta
    CLF Andrea COM-B3S-A3S-buy one dog but NEG A1S-know-PRF NEG
   ankuchi ri'.
   where DET
   Intended: 'Andrea bought a dog, but I don't know where it was.'

Third, spading is incompatible with mention-some modification, in contrast to the SLCs discussed earlier:

(67) *Spading in Kaqchikel and mention-some-modification*
  a. A: Ta-loq'-o jun spanïk chin a-te'!
     IMP-buy-SS a gift RN A2S-mother
     'Buy a gift for your mother!'
    B: Achike ta n-Ø-in-löq'?
     what IRR INC-B3S-A1S-buy
     'What do I buy for example?'

b.　　A: Ta-loq'-o　　jun spanïk chin a-te'!
　　　　　IMP-buy-SS a　　gift　　RN A2S-mother
　　　　　'Buy a gift for your mother!'
　　　　B: *Achike ta　k'a　ri'?
　　　　　what　　IRR PART DET
　　　　　Intended: 'What for example is it?'

c.　　A: Ta-loq'-o　　jun spanïk chin a-te'!
　　　　　IMP-buy-SS a　　gift　　RN A2S-mother
　　　　　'Buy a gift for your mother!'
　　　　B: *Achike ta　jun　ri'?
　　　　　what　　IRR one DET
　　　　　Intended: 'What for example is it?'

Fourth, spading is incompatible with else-modification, in contrast to SLCs:

(68)　*Spading in Kaqchikel and else-modification*

　　　　A: Ma Juan x-Ø-u-löq'　　　äk'.
　　　　　CLF Juan COM-B3S-A3S-buy chicken
　　　　　'Juan bought chicken.'

　　　　B: *Aw-etama-n　achike chmas ri'?
　　　　　A2S-know-PRF what　more　DET
　　　　　Intended: 'Do you know what else it was?'

To summarize: All four diagnostics show that spading behaves like a reduced cleft/pseudosluice. A table summarizing the diagnostics applied to SLCs and spading is given below:

(69)　*Ellipsis vs. reduced cleft diagnostics*

	KAQCHIKEL CONSTRUCTION			
DIAGNOSTIC	*SLC (wh-remnant)*		*Spading (wh-remnant+determiner)*	
allows agg. non-D-linked *wh*	NO	(47)-(49)	YES	(65)
allows sprouting	YES	(53)-(58)	NO	(66)
allows mention some modification	YES	(60)	NO	(67)
allows else modification	YES	(62)	NO	(68)
CONCLUSION	*patterns with sluicing*		*patterns with pseudosluicing*	

I conclude that SLCs instantiate true sluicing, not pseudosluicing. Any proposal that would take SLCs to instantiate pseudosluicing would face not only the challenges raised by the results of these four diagnostics as applied to SLCs, but it would also be forced to explain why there is an independent construction in Kaqchikel (spading) that patterns with pseudosluicing in the same respects. I will therefore consider SLCs to be sluices moving forward and will refer to the data as sluicing.

2.2.3 Voice mismatches in Kaqchikel sluicing

This section explores possible and impossible voice mismatches in Kaqchikel sluicing. Before delving into the data, let us emphasize how we will determine the voice specification within the sluice.

As we saw in the previous sections, some voices are incompatible with A'-extraction of certain elements. First, recall the following: When the external argument of a transitive is extracted, active voice is impossible, and AF is required (note that I use < > here to notate a lower copy of a moved *wh*-phrase>:

(70)   *Extracting the EA in Kaqchikel; AF required*
         a. * *wh*-EA V-ACT IA <*wh*-EA>
         b. ✓*wh*-EA V-**AF**   IA <*wh*-EA>

We will use this property of Kaqchikel grammar in the following way: In Kaqchikel sluicing, if the *wh*-remnant is the Agent of a dyadic predicate, and that *wh*-remnant is not encased in a relational noun structure, then the voice specification within the sluice must be AF (though note that it seems to be impossible to ask a *wh*-question of the thematic

agent of a passive via a *wh*-question encased in a relational noun structure; see section 2.5).[26]

Similarly, we observed that object and adjunct *wh*-extraction is banned with AF.

(71)   *Extracting the IA in Kaqchikel; AF banned*
    a. * *wh*-IA V-**AF**   <*wh*-IA> EA
    b. ✓ *wh*-IA V-ACT <*wh*-IA> EA

(72)   *Extracting an adjunct in Kaqchikel; AF banned*
    a. * *wh*-adjunct V-**AF**   IA EA <*wh*-adjunct>
    b.   *wh*-adjunct V-ACT IA EA <*wh*-adjunct>

Again, we will use this particular aspect of the grammar of Kaqchikel in the following way: In Kaqchikel sluicing, if the *wh*-remnant is an internal argument (IA) or an adjunct, then the voice specification within the sluice cannot be AF.

Finally, an IA cannot be A'-extracted in the AP voice:

(73)   *Extracting the IA in Kaqchikel; AP banned*
    a. * *wh*-IA V-**AP**   <*wh*-IA> EA
    b. ✓ *wh*-IA V-ACT <*wh*-IA> EA

In a nutshell then, the grammar of Kaqchikel allows us to manipulate voice mismatches in a controlled way through the form of the *wh*-remnant in sluicing. Since the language's grammar forbids the use of certain voices with A'-extraction of specific elements, we can diagnose with precision what voice specification must be present in the sluice. The

---

[26] The APOBL is also possible here, but I will not discuss that voice until section 2.5.1. As we will see, however, it will be irrelevant whether the sluice is APOBL or AF, since I will argue that these two constructions have the same syntax and lack Voice^0 altogether. Furthermore, we will see that independent considerations will suggest that APOBL cannot be in the ellipsis site when the antecedent is active or passive (Chung 2006's constraint on "new words" in the ellipsis site; see chapter 5 of the present work).

possibility that the constraints in question are themselves island constraints, and may be nullified under ellipsis (so-called repair-by-ellipsis), has been taken into account in the analysis; see 2.4 for details.

A proviso is necessary before moving forward. We established before that there are five voices in Kaqchikel, so there are 20 logically possible voice combinations that we could investigate. In the next few subsections, I only lay out the most relevant mismatches in order to be able to formulate an analysis in section 2.3, returning to the full range of possible combinations in section 2.5.

Let us move on, then, to voice mismatches in Kaqchikel sluicing.

2.2.3.1 AP-Active mismatches

The first mismatch we will discuss involves an antecedent with AP voice and a sluice containing active voice. Extracting the IA in an AP (or AF) clause is impossible, as shown in (74)b. To extract an IA, the active voice must be used:

(74)   *AP with IA extraction*
         a. Y-e-tik-on.
            INC-B3P-plant-AP
            'They plant (corn).'

         b. *Achike ixim y-e-tik-on?
              what   corn INC-B3P-plant-AP
              Intended: 'What corn do they plant?'[27]

         c. Achike ixim n-Ø-ki-tïk?
            what   corn INC-B3S-A3P-plant
            'What corn do they plant?'

---

[27] I use 'which corn' here in order to make clear that the interpretation that is being tested involves IA extraction. If a bare *wh*-word *achike* were used instead, a confound would be introduced: *Achike* also means 'who', so the sentence would be licit, but under an interpretation were the EA is extracted, rather than the IA.

Given the data above, we can force a voice mismatch under sluicing by having an IA *wh-*remnant. Using this manipulation, the example below shows that it is not possible to mismatch the AP voice with active voice under sluicing. Note that the data in (75) reflects two question-answer pairs (i.e., dialogues):

(75)   *AP-Active voice mismatch*

A: Y-e-tik-on.
   INC-B3P-plant-AP
   'They are planting corn.'

B: *Achike ixim <n-Ø-ki-tïk>?
   what   corn   INC-B3S-A3S-plant
   Intended: 'What corn?'

B′: Achike ixim n-Ø-ki-tïk,   säq, xwana,   o q'än?
    what   corn INC-B3S-A3P-plant white black.corn or yellow
    'What corn do they plant, white, black, or yellow?'

(76)   *AP-Active voice mismatch*

a. Yïn x-i-loq'-on=pe.   pa k'ayib'äl. Ta-wla   achike
   1S COM-B1S-buy-AP=DIR PREP market   IMP-guess what
   x-Ø-in-löq'=pe!
   COM-B3S-A1S-buy=DIR
   'I bought (something) at the market. Guess what I bought!'

b. *Yïn x-i-loq'-on=pe   pa k'ayib'äl. Ta-wla   achike
    1S   COM-B1S-buy-AP=DIR PREP market   IMP-guess what
    <x-Ø-in-löq'=pe>!
     COM-B3S-A1S-buy=DIR
    Intended: 'I bought at the market. Guess what!'

(77)   *AP-Active voice mismatch*

a. Xta Juana n-Ø-k'ay-in   pa k'ayib'äl wkami. Ta-wla   achike
   CLF Juan INC-B3S-sell-AP PREP market   now   IMP-guess what
   n-Ø-Ø-k'ay-ij!
   COM-B3S-A3S-buy-ACT
   'Juana is selling (something) at the market now. Guess what she is selling!'

b. *Xta Juana n-Ø-k'ay-in   pa k'ayib'äl wkami. Ta-wla   achike
    CLF Juan INC-B3S-sell-AP PREP market   now   IMP-guess what
    <n-Ø-Ø-k'ay-ij>!
      COM-B3S-A3S-buy-ACT
    Intended: 'Juana is selling (something) at the market now. Guess what!'

The takeaway of this subsection is the following: Kaqchikel disallows mismatching the AP voice in the antecedent with an active voice in the ellipsis site under sluicing. This fact is important, because the generalization that will arise from the Kaqchikel data is that *some*, but not all, voice mismatches are well-formed.[28]

We can already be sure, then, that it is not the case that Kaqchikel allows voice mismatches freely. In other words, we could not propose that there is complete freedom in Kaqchikel regarding ellipsis, with no identity condition regulating silence in discourse. Rather, there are types of mismatches that are impossible and others that are possible, as we will see and is expected. It is necessary, then, to formulate an identity condition that can cover the totality of the data, in Kaqchikel and more broadly cross-linguistically (i.e., the English data discussed earlier).

2.2.3.2 Active-AF mismatches

Here, I will show that it is possible to mismatch active and AF under sluicing. To be more precise, certain configurations *force* Active-AF mismatches due to particularities of Kaqchikel grammar, and the result is nevertheless well-formed.

Recall first that *wh*-movement of the EA requires the use of AF. As a result, if we construct examples where the *wh*-remnant is an EA (an Agent argument not enclosed in a relational-noun structure), then we can be confident that the content of the sluice will not

---

[28] A question that arises here is why a derivation is not available for the data in (75)-(77) where there is AP voice in the ellipsis site and whatever problem exists with extracting the theme of an antipassive clause is repaired by sluicing. I leave this issue for future research, but note that we need to assume that the issue with the relevant extraction in the antipassive lies in the syntactic component, rather than at PF (see Mendes 2020 for discussion on salvation and non-salvation by deletion formulated along these lines; see chapter 4 here).

be in the active voice. Instead, the contents of the ellipsis site must be an AF clause. An example of this sort is observed below:

(78)　*Active-AF mismatch; EA wh-remnant*

    a. Jun winäq　x-Ø-Ø-tzak-**ij**　　　la　che'... Man　w-etama-n　　　ta
       one person COM-A3S-E3S-**cut**-ACT DEM tree　NEG　GEN1S-know-PRF NEG
       achike winäq < x-Ø-tzak-**in**　　　la　che' >.
       which　person　COM-A3S-cut-**AF** DEM tree
       'A person cut that tree... I don't know which person.'

    b. Yïn x-Ø-in-tz'ët　　　chi　jun ixöq　x-Ø-u-chöy　　　la　che' la'. Man
       1S　COM-B3S-A1S-see COMP one woman COM-B3S-A3S-cut DEM tree　DEM NEG
       x-Ø-in-tz'ët　　　ta　jab'ël achike ixöq　<x-Ø-choy-o　　la　che' la'>.
       COM-B3S-A1S-see NEG well　what　woman　COM-B3S-cut-AF DEM tree　DEM
       'I saw that a woman cut that tree. I didn't see clearly which woman.'

The examples above instantiate sluices where the voice specification of antecedent and ellipsis site mismatch—while the antecedent is in the active voice, the presence of an EA *wh*-remnant ensures that the ellipsis site is not in the active voice, but instead contains AF.[29]

2.2.3.3 AF-Active mismatches

Here, we will consider examples that instantiate an AF-Active mismatch; i.e., the converse of the mismatch discussed in the preceding subsection. Once again, we will make use of certain configurations in the language to force the voice mismatch.

Recall first that *wh*-movement of the IA is incompatible with the use of AF. Instead, active voice is required in such a configuration. Consider, then, the example below, where the antecedent is an AF clause. Given that the *wh*-remnant in the target clause is an IA, we

---

[29] In section 2.3.4 an alternative analysis will be assessed, where the sluice contains active voice and the example is licit due to island repair by ellipsis. We will observe that this is a feasible analysis only for this type of mismatch, but not the others, and can therefore be rejected on the grounds that it will not generalize.

can be sure that the ellipsis site does not contain AF, but instead must be in the active voice.

Nevertheless, as shown in (78)B″, this is a well-formed instance of sluicing:

(79)   *AF-Active; IA wh-remnant*

    A: Xa   xe   ri   ma   Juan x-Ø-loq'-o   kotz'i'j.
        EMPH only DET CLF Juan COM-B3S-buy-AF flower
        'Only Juan bought flowers.'

    B: Kan qitzij? Ta-b'ij   pe   chwe   achike kotz'i'j x-Ø-u-löq'!
        INT truth   IMP-say DIR PREP.A1S.RN what   flower COM-B3S-A3S-buy
        'Really? Tell me which flowers he bought!'

    B′: *Kan qitzij? Ta-b'ij   pe   chwe   achike kotz'i'j x-Ø-loq'-o!
         INT truth   IMP-say DIR PREP.A1S.RN what   flower COM-B3S-buy-AF
        Intended: 'Really? Tell me which flowers he bought!'

    B″: Kan qitzij? Ta-b'ij   pe   chwe   achike kotz'i'j <x-Ø-u-löq'>!
         INT truth   IMP-say DIR PREP.A1S.RN what   flower   COM-B3S-A3S-buy
        'Really? Tell me which flowers!'

There is another configuration that we can use to provide evidence that AF-Active mismatches are licit in Kaqchikel. Recall that *wh*-extraction of adjuncts is incompatible with AF. Consider then the examples below, where the antecedent is in the AF voice and the target clause contains a sprouted adjunct. Given that the *wh*-remnant is an adjunct, we can be sure that the ellipsis site does not contain AF and must instead be in the active voice:

(80)   *AF-Active; reason adjunct sprouting*

    a. K'o   jun x-Ø-loq'-o   ri   aq. Aw-etama-n   achike ru-ma
        EXIST one COM-B3S-buy-AF DET pig A2S-know-PERF what   A3S-RN
        x-Ø-u-löq'?
        COM-B3S-A3S-buy
        'Someone bought the pig. Do you know why he/she bought it?'

    b. *K'o   jun x-Ø-loq'-o   ri   aq. Aw-etama-n   achike ru-ma
        EXIST one COM-B3S-buy-AF DET pig A2S-know-PERF what   A3S-RN
        x-Ø-loq'-o?
        COM-B3S-buy-AF
        Intended: 'Someone bought the pig. Do you know why he/she bought it?

    c. K'o   jun x-Ø-loq'-o   ri   aq. Aw-etama-n   achike ru-ma
        EXIST one COM-B3S-buy-AF DET pig A2S-know-PERF what   A3S-RN

<x-Ø-u-löq'>?
  COM-B3S-A3S-buy
'Someone bought the pig. Do you know why?'

(81) *AF-Active; locative adjunct sprouting*

    a. Xa   xe   ri   ma Pedro x-Ø-loq'-o   ri   kotz'i'j. Aw-etaman
      EMPH only DET CLF Pedro COM-B3S-buy-AF DET flower A2S-know-PERF
      ankuchi x-Ø-u-löq'   wi?
      where   COM-B3S-A3S-buy-AF FP
      'Only Pedro bought the flowers. Do you know where he bought them?'
    b. Xa   xe   ri   ma Pedro x-Ø-loq'-o   ri   kotz'i'j. Aw-etaman   ankuchi
      EMPH only DET CLF Pedro COM-B3S-buy-AF DET flower A2S-know-PERF where
      <x-Ø-u-löq'   wi>?
      COM-B3S-A3S-buy FP
      'Only Pedro bought the flowers. Do you know where?'

The examples we have discussed here show that AF-Active mismatches are licit in Kaqchikel.

2.2.3.4 Passive-AF mismatches

In this subsection, I will show that the grammar of Kaqchikel forces Passive-AF mismatches under sluicing in certain configurations as well.

    Consider first the sluicing example below. Here, the antecedent is in the passive voice and the demoted agent is an indefinite introduced by the relational noun *uma*. In this dialogue, speaker B uses a sluice with an EA *wh*-remnant whose correlate is the indefinite thematic agent in the antecedent. Like in previous discussions, the EA *wh*-remnant ensures that we are dealing with AF in the sluice.

(82) *Passive-AF mismatch*

    A: Ri aq x-Ø-kam-is-äx   r-uma   jun ixöq.
       DET pig COM-B3S-die-CAUS-PAS A3S-RN   a   woman
       'The pig was killed by a woman.'

    B′: Kan qitzij? Achike ixöq   <x-Ø-kam-sa-n   ri   aq>?
       INT truth which woman COM-B3S-die-CAUS-AF DET pig
       'Really? Which woman?'

B″: Kan qitzij? Ta-b'ij pe   chwe   achike ixöq   < x-Ø-kam-sa-n
    INT truth   IMP-say DIR PREP.A1S.RN which woman   COM-B3S-die-CAUS-AF
    ri   aq>!
    DET pig
    'Really? Tell me which woman!'

The examples below show that non-sluice variants of B's utterance above are ill-formed

unless AF is used.

(83)   *Non-elliptical version of (82)B*

    a. Kan qitzij? Achike ixöq   x-Ø-kam-sa-n?
      INT truth   which woman COM-B3S-die-CAUS-AF
      'Really? Which woman killed it?'

    b. *Kan qitzij? Achike ixöq   x-Ø-u-kam-sa-j?
      INT truth   which woman COM-B3S-A3S-die-CAUS-ACT
      Intended: 'Really? Which woman killed it?'

    c. *Kan qitzij? Achike ixöq   x-Ø-kam-is-äx?
      INT truth   which woman COM-B3S-die-CAUS-PAS
      Intended: 'Really? Which woman was it killed by?'

We therefore observe that Passive-AF mismatches are also allowed in Kaqchikel.

2.2.3.5 Summary of voice mismatches (so far)

The table below summarizes the mismatches we discussed in this section. I reiterate that

these are not all possible mismatches one could conceivably construct, but they will suffice

for now in order to lay out an analysis.

(84)   *Voice mismatches in Kaqchikel sluicing (version 1)*

Antecedent	*Sluice*	*Status*	*Example*
AP	Active	*	(75)-(76)
Active	AF	✓	(78)
AF	Active	✓	(79)-(81)
Passive	AF	✓	(82)

Before transitioning to the next section, it must be emphasized that AF has to be analyzed

as a member of the voice system in Kaqchikel. In other words, the data we have discussed

in this chapter could not be accounted for by proposing that AF can "mismatch" with active

75

and passive in sluicing because AF is not a part of the voice system at all—i.e., by positing that the AF exponent is a morphological quirk of some kind. To illustrate the gist of such an alternative, notice that in an example like the one below, we would not claim that active voice can mismatch with *plural*—these features are hosted by different heads, so no conclusions should be drawn about the identity condition from the well-formed status of this kind of configuration:

(85)   Beth came to the party, but I don't know **which other representatives** did.

However, claiming that a state of affairs like in (85) applies to the voice mismatches involving AF summarized in (84) runs into an unsurmountable problem. In brief, one would need to propose that AF clauses are formally *active*, in order to rule in the Active-AF and AF-Active mismatches, but one would also need to propose that AF clauses are formally *passive*, in order to rule in the Passive-AF mismatches. Crucially, even though Active-Passive mismatches in sluicing cannot be tested in Kaqchikel (see section 2.5), we know that this mismatch is ill-formed cross-linguistically. Thus, there is little hope for an alternative that would seek to exclude AF from the voice system in the language and use such a proposal to account for the data discussed thus far.

We also discussed in chapter 1 the existence of approaches to ellipsis that minimize the work done by an identity condition and prioritize instead the information-structural status of ellipsis remnants as the explanatory tool to account for contrasts between well-formed and ill-formed mismatches (e.g., Kertz 2013). It is unclear, however, how an approach of this sort could explain the English and Kaqchikel patterns. All of the sluicing examples by definition involve a *wh*-remnant in the target clause, whose information-structural status is of a similar nature. Nevertheless, a subset of voice mismatches is ill-formed (Active-

Passive, Passive-Active, AP-Active), while another is well-formed (Active-AF, AF-Active, Passive-AF; see section 2.5.1 for evidence that AP$_{OBL}$-Active mismatches are also allowed). How could pragmatic considerations deliver the empirical picture we laid out? I can conceive of no explanatory account under this kind of approach, but encourage work in those frameworks to take into consideration the results presented here.

With this discussion in mind, we are now in a position to analyze why a subset of voice mismatches are well-formed in Kaqchikel sluicing, in contrast to the findings in the literature on voice mismatches in languages like English.

2.3 Featural non-distinctness as an identity condition on ellipsis

In this section, I provide an analysis of AF and propose a novel syntactic identity condition for ellipsis. I depart from previous analyses of AF (e.g. Ordóñez 1995, Stiebels 2006, Aissen 2011, Coon et al. 2014, Assman et al. 2015, Erlewine 2016, Deal 2016, Coon et al. 2021) and propose that the construction instantiates a configuration in which the Voice layer has been removed via Exfoliation, a Last Resort operation triggered under particular structural conditions (Pesetsky 2021). The proposed analysis explains why Active-AF, AF-Active, and Passive-AF mismatches are allowed in Kaqchikel: none of these mismatches instantiate a clash between a pair of fully specified feature sets. Instead, the mismatch involves a featurally specified Voice head (e.g., Voice$_{[F:ACTIVE]}$, which I notate interchangeably as Voice$_{ACTIVE}$) in opposition with the absence of a Voice head (AF). I propose that this type of mismatch, where featural *non-distinctness* is obeyed, is always allowed under ellipsis (in Kaqchikel and cross-linguistically). This approach to syntactic identity has desirable consequences, since it sheds light on why certain familiar mismatches outside of the voice domain are allowed (e.g., tense mismatches), as we will delve into in

77

detail in chapters 3 and 4 of this dissertation. Additionally, the analysis provides an explanation of many of the features specific to AF that have received attention in the literature, including (i) the Last Resort distribution of AF, (ii) the unavailability of ERG agreement in AF clauses, and (iii) the ability of ABS agreement to target either the external or the internal argument in AF. I will discuss these properties in section 2.4.

2.3.1 Revisiting the empirical puzzle

Recall the central data that have driven this investigation. Whereas a mismatch between active and passive clauses has been observed to be impossible in sluicing cross-linguistically, certain voice mismatches are allowed in Kaqchikel. An active clause can serve as the antecedent for a sluice in the AF voice and vice-versa. We also observed that Passive-AF mismatches are possible, though I do not repeat those data here:

(86)　　*Active antecedent-AF sluice*
　　　　a. Jun winäq　x-Ø-Ø-tzak-**ij**　　　　la　che'... Man w-etama-n　　ta
　　　　　 one person COM-B3S-A3S-**cut**-ACT DEM tree　NEG A1S-know-PRF NEG
　　　　　 achike' winäq < x-Ø-tzak-**in**　　　la　che'>.
　　　　　 which　person　COM-B3S-cut-**AF** DEM tree
　　　　　 'A person cut that tree... I don't know which person.'
　　　　b. X-Ø-u-löq'　　　 jun monton kotz'i'j jun winäq, po　man w-etama-n　　　ta
　　　　　 COM-B3S-A3S-buy one bunch　flower one person but　NEG A1S-know-PERF NEG
　　　　　 achike winäq <x-Ø-loq'-**o**　　　 jun monton kotz'i'j>.
　　　　　 which　person COM-B3S-buy-**AF** one bunch　flowers
　　　　　 'Some person bought a bunch of flowers, but I don't know which person.'

(87)　　*AF antecedent-Active sluice*
　　　　a. Ja　ri　xta Maria x-Ø-kam-sa-**n**　　　　jun　chiköp… Pero man
　　　　　 FOC DET CLF Maria COM-B3S-die-CAUS-**AF** one　animal　but　NEG
　　　　　 w-etama-n　　　ta　achike chiköp <x-Ø-u-kam-sa-**j**　　　　　　ri
　　　　　 GEN1S-know-PRF NEG which animal　COM-A3S-E3S-die-CAUS-**ACT** DET
　　　　　 xta　Maria>.
　　　　　 CLF Maria
　　　　　 'MARIA killed an animal... but I don't know which animal.'
　　　　b. Ja　yïn x-i-kam-sa-**n**　　　　 jun chiköp. Aw-etama-n　　achike chiköp
　　　　　 FOC 1S　COM-B1S-die-CAUS-**AF** one animal　A2S-know-PERF which　animal

        <x-Ø-in-kam-sa-**j**   yïn>?
        COM-B3S-A1S-die-CAUS-**ACT** 1S
        'I killed an animal. Do you know which animal?'

The observation that voice mismatches are banned in sluicing has led to approaches arguing that some syntactic identity must be obeyed under ellipsis. Merchant (2013) proposes a condition along the following lines:

(88)   *Merchant (2013)'s syntactic identity condition (repeated from (4))*

       The heads in the verbal spine of the elided constituent must be syntactically identical to the corresponding heads in the antecedent.

As should be apparent from our discussion of Kaqchikel, conditions of this nature are too strong—they undergenerate, predicting that the Kaqchikel data should not be well-formed. I will argue instead that what matters in the calculation of syntactic identity in ellipsis is not *absolute* identity, as Merchant proposes. Rather, the relevant notion in the calculation of syntactic identity is featural non-distinctness. Below, I provide the identity condition that I will defend here and in the rest of the dissertation—in chapter 4, however, the condition will be modified to include a strict √ROOT identity requirement:

(89)   *Syntactic identity in ellipsis (repeated from (1))*

       The antecedent and material properly contained within the ellipsis site must be featurally non-distinct.

One aspect of the identity condition that will not play a role in this chapter, but will become crucial in chapters 3-4, is the specification that identity is calculated only over material that is *properly* contained in the ellipsis site.

The condition in (89) has its roots in Chomsky 1965: 177-182, which argued that deletion under identity in comparatives satisfies non-distinctness, as opposed to strict identity. For example, consider '*I know several more successful lawyers than Bill*.', where a plural feature in the antecedent does not clash with singular, which is assumed to be

unspecified for number in the base structure (Chomsky 1965: 181; see Lipták 2013 and Lasnik & Funakoshi 2019). Chomsky, however, does not show that non-distinctness can be bidirectional, in contrast to what is required by the Kaqchikel data discussed here.

Let us discuss, then, the consequences of this condition for our purposes in this chapter: accounting for the range of voice mismatches that are well- and ill-formed in sluicing. I am adopting the view that there is a clausal spine (a sequence of projections) and features are distributed across those heads. I also assume that the difference between active and passive voice lies in the value of the attribute for the feature borne by the Voice head. Given this assumption, the different featural specifications given below for the Voice head in the antecedent and the Voice head in the sluice result in a full featural clash, thus violating the condition in (89):

(90)   *Voice mismatch under sluicing—full featural clash banned*
        <u>Antecedent</u>   <u>Sluice</u>
        a. *Voice[F:ACT]   Voice[F:PASS]   *English*
        b. *Voice[F:PASS]   Voice[F:ACT]   *English*
        c. * Voice[F:ANTIPASSIVE]   Voice[F:ACT]   *Kaqchikel*

Note that there is a degree of freedom in the featural geometry that one could assume for Voice, in English or in other languages—for example, one could instead adopt a proposal where the distinction between active and passive voice lies in a different privative feature borne by Voice (Voice[ACTIVE] and Voice[PASSIVE]). For our purposes here, choosing between this alternative featural representation for Voice and the one in (90) does not deliver a different result—I will assume that such a configuration also violates featural non-distinctness (see below). Nevertheless, it would be a productive future extension of the work here to explore whether a different conception of what does or does not count as

80

non-distinct could shed light on how the feature geometry of Voice (and other heads) should be represented.

Returning to the identity condition proposed here, the mismatches below are predicted to be well-formed, in contrast to those in (90). In the following sluices, either the sluice (91)a,c or antecedent (91)b lacks a Voice head altogether—i.e., either the antecedent or sluice is a structural subset of the other. Therefore, mismatching is allowed, since these instances of ellipsis comply with the condition in (89). Antecedent and sluice are non-distinct and thus licit:

(91)   *Voice mismatch under sluicing—featural non-distinctness allowed*

Antecedent	Sluice	
a. Voice$_{[F: ACT]}$	Ø (AF)	*Kaqchikel*
b. Ø (AF)	Voice$_{[F: ACT]}$	*Kaqchikel*
c. Voice$_{[F: PASS]}$	Ø (AF)	*Kaqchikel*

In a nutshell, mismatches are allowable as long as the mismatch does not involve a feature clash.[30]

In chapter 3 and 4, another configuration will be assessed that also satisfies featural non-distinctness. Specifically, I will discuss cases where antecedent and ellipsis site contain a head H, but the identity condition is satisfied because one of the instances of H does not bear a feature at all (e.g., the antecedent contains H$_{[F]}$ and the ellipsis site H$_{Ø}$). In

---

[30] Note that disallowable ellipses like the following, involving the presence of lexical material like *pieces* in the ellipsis that is not present in the antecedent, fall outside of the scope of the condition thus far:

(i)   *Kim ate, but I don't know of what <Kim ate pieces>.

These data are ruled out by Chung (2006)'s lexical condition. We will discuss data of this type in chapter 4, where we will defend an addendum to (89) revolving around strict √ROOT identity (Saab 2008); see chapter 5 for discussion.

such a case, there is no featural clash either and the configuration satisfies featural non-distinctness. Finally, chapter 3 will touch upon cases related to [PERSON] mismatches where the antecedent and the ellipsis site contain a head H, but each instance of H bears a feature that is not of the same nature as the other (e.g., the antecedent contains $H_{[F]}$ and the ellipsis site $H_{[G]}$; see above on the alternative configuration for Voice where active and passive each bear a different privative feature). I will assume that this configuration does create a featural clash, violating the identity condition, which requires non-distinctness.

Let us now analyze concretely how AF clause lacks Voice^0. I will show that the previously discussed mismatches instantiate the configurations in (91), thus complying with the syntactic identity condition in (89).

2.3.2 AF as Voice Exfoliation

The structure I assume for regular active transitive clauses in Kaqchikel is the following:[31]

---

[31] I do not delve into the structure of intransitive clauses. I assume that in intransitives, a different Voice$_{ACT}$ whose probing does not result in ERG agreement is merged. Note that it is independently necessary for different Voice heads to be able to select for distinct flavors of v. For example, one needs to specify that Voice$_{PASS}$ cannot select for an intransitive v in languages that do not allow intransitive verbs to passivize.

(92) *Kaqchikel active transitive clause*

```
CP
  \
   C'
     \
      InflP
         \
          Infl'
            /    \
    Infl[ABS]    VoiceP        ←Phase
         \      /      \
          `-- IA       Voice'
                      /      \
          Voice_ACT[ERG, EPP]   vP
                   \        /    \
                    `----. EA     VP
                                 /   \
                                V     IA
```
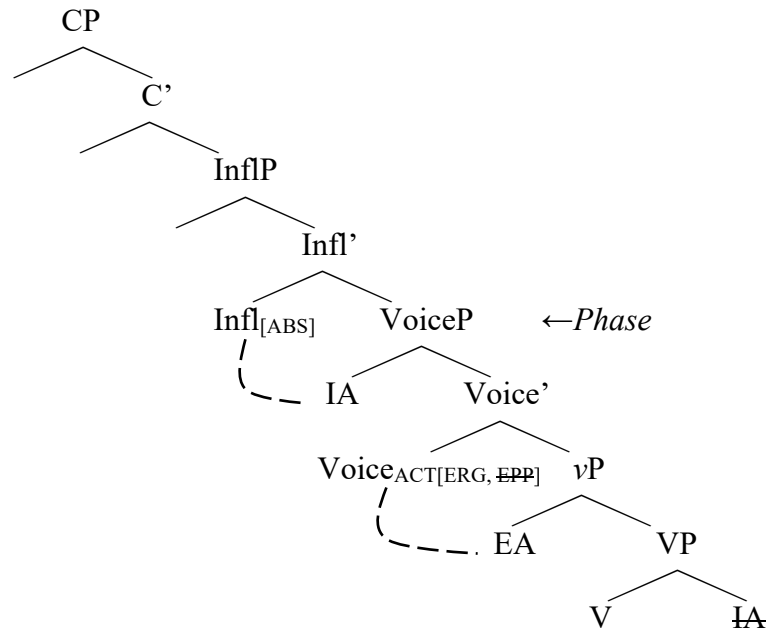
I assume that the verbal domain contains a Voice and a v layer, with Voice^0 above v^0 (Legate 2014; see Harley 2017 on splitting Voice^0 and v^0; see also Kratzer 1996). Voice^0, the topmost layer, is a phase (for phase theory, see Chomsky 2001, Citko 2014), while v^0 introduces the external argument (Merchant 2013, Coon et al. 2014; see Collins 2005). The EA enters into an Agree relation with the active Voice head, resulting in a doubled clitic on Voice that surfaces as ergative (Set A) agreement on the verbal stem (Preminger 2014).[32] The internal argument moves to the specifier of the Voice projection—the phase-

---

[32] I depart from proposals that assume that ERG agreement is the result of an inherent relation between $v_{[active]}$ and the EA introduced in Spec,vP (e.g., Coon 2017 for Ch'ol). I do not think that there are any consequences of assuming that ERG is the result of probing by Voice instead of inherent agreement with the head that introduces the EA (although, see Baker 2015 chapter 2 for some comments on general issues arising from ERG Case being assigned through Agree by a lexical head such as T^0). There seems to be little evidence for an inherent case approach to ERG in Kaqchikel, though many scholars assume this position (see Baker &

edge—and is therefore accessible to probing by $Infl^0$ (Coon et al. 2014 for Q'anjob'al, Douglas et al. 2017 for Mayan more generally; see also Aldridge 2004 for Austronesian). $Infl^0$ agrees with the internal argument, resulting in a doubled clitic on $Infl^0$ that surfaces as absolutive (Set B) agreement. Note that the probing of $Voice_{ACT}$ which results in clitic doubling of the EA needs to be dissociated from the probing by the same head which targets the IA and raises it to Spec,VoiceP, satisfying the EPP feature.[33] I assume that when $Voice^0$ clitic doubles the EA (thus creating a chain), the EA is no longer an intervener for EPP related probing (see Anagnostopoulou 2003 for evidence that clitic-doubled nominals are not interveners). The IA is therefore the nominal which satisfies the EPP on Voice.

Crucially, it is the configuration in (93) that gives rise to the extraction restriction on the subject of transitive clauses in Kaqchikel (see references above). In concrete terms, the extraction restriction which bans A'-movement of the EA in an active clause is a result of the clausal configuration in Kaqchikel transitive clauses and the Phase Impenetrability Condition (PIC; Chomsky 2001), which renders elements in the complement of a phase head unavailable to syntactic operations such as probing by Agree. For our purposes here, it is inconsequential whether one adopts the so-called PIC1 (Chomsky 2000) or PIC2

Bobaljik 2017 for a comparison of inherent vs. dependent case approaches to languages with morphological case on nominals). As we will see, the configuration I propose has desirable consequences in explaining the particularities of AF regarding, among other things, the observed agreement patterns.

[33] In other words, Voice must probe twice: the first probing finds the EA, while the second probing raises the IA for EPP satisfaction. I leave for future work the question of whether a deep fact about the grammar derives the existence of only this particular ordering of probing (as opposed to the inverse ordering), or whether the ordering is a parameter of sorts. If the latter is correct, we would make predictions about other possible grammars.

(Chomsky 2001). What is crucial is that only the element at the edge of the phase is accessible and VoiceP in Kaqchikel is restricted to a single specifier: since the IA occupies such a position, the EA is trapped, inaccessible to probing by C^0 and unable to move. An illustration of the issue is given below.[34]

(93)   *Extraction restriction in active transitive clause*
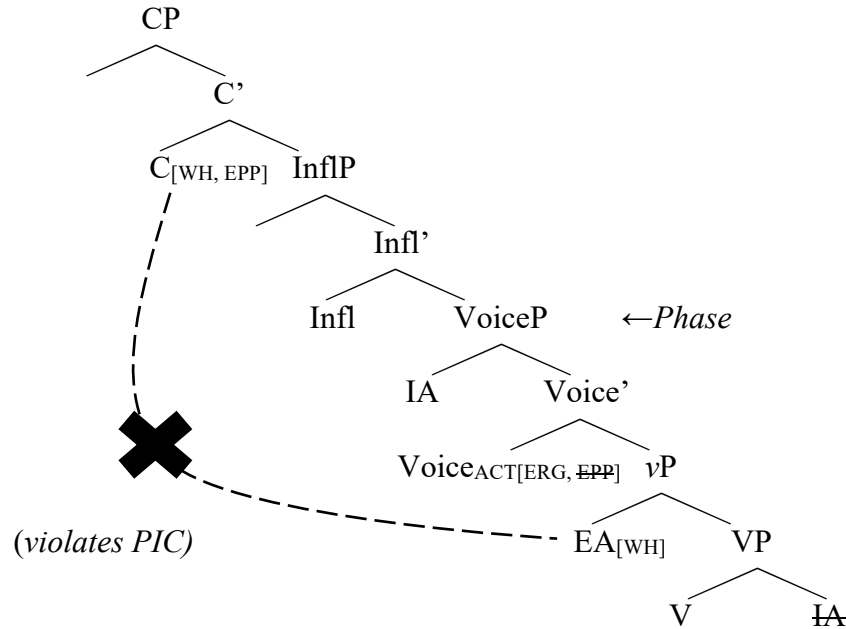
The way to circumvent the extraction restriction is AF.[35] There are two ways in which one could envision AF circumventing the extraction restriction. First, one could assume that AF is a special type of Voice, specified for a set of features, which allows for the extraction of the EA somehow. Below I sketch a possible analysis of this ilk (see Coon et al. 2014, Assman et al. 2015)—here, Voice$_{AF}$ extraordinarily allows for the EA to extract:

---

[34] Assman et al. (2015) refer to this analysis as the "absolutive island" approach, by analogy with a *wh*-island. This is because movement of the absolutive IA to the edge of the phase closes off the domain, creating an island for extraction.

[35] The oblique antipassive can also be used. I will discuss the oblique antipassive in section 2.5.

(94)　*AF circumvents extraction restriction*

```
                    CP
                  /    \
               EA       C'
                      /    \
              C[WH, EPP]    InflP
                  |        /    \
                  |      Infl'
                  |     /    \
                  |  Infl[ABS]   VoiceP        ←Phase
                  |            /      \
                  |         EA         Voice'
                  |                   /      \
                  |              Voice_AF     vP
                  |                          /   \
                  ⌐ ⋯ ⋯ ⋯ ⋯ EA[WH]    VP
                                         /    \
                                        V      IA
```
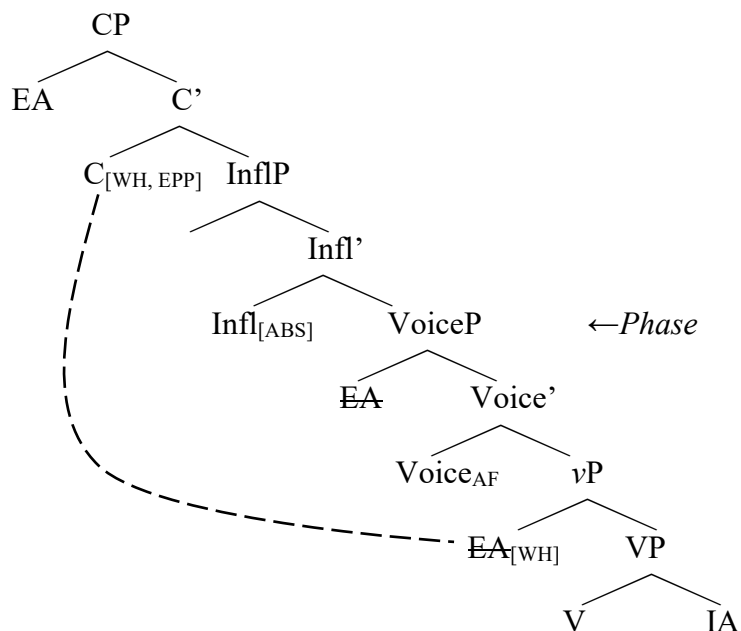
Taking a viewpoint on AF as sketched in (94) faces the problem of how to rule in the attested voice mismatches under sluicing. If AF is a fully specified Voice$_{[AF]}$, with a featural specification that differs from the featural specification of Voice$_{[ACT]}$ and Voice$_{[PASS]}$, then we would predict that a mismatch under sluicing should result in ungrammaticality, parallel in kind to disallowed passive and active mismatches in English. An analysis like (94) would thus require a specifically-tailored proposal for how Voice$_{AF}$ is a special Voice^0 that falls outside the domain of the identity condition on ellipsis (for an analysis of AF in this spirit, see Coon et al. 2021: section 5, which proposes that AF involves a v_{AF} head).

Instead of taking this viewpoint on AF—which could not derive the sluicing facts discussed in this chapter—I propose a novel approach to the construction which has the potential to illuminate the reason behind the availability of voice mismatches, while also explaining many of the idiosyncratic properties of AF itself in the process (see 2.5). I

propose that AF is the result of removing the Voice layer in the course of the derivation. Before delving into the specific details of the proposal, an illustrative sketch is provided below:

(95)   *AF instantiates the absence of Voice*
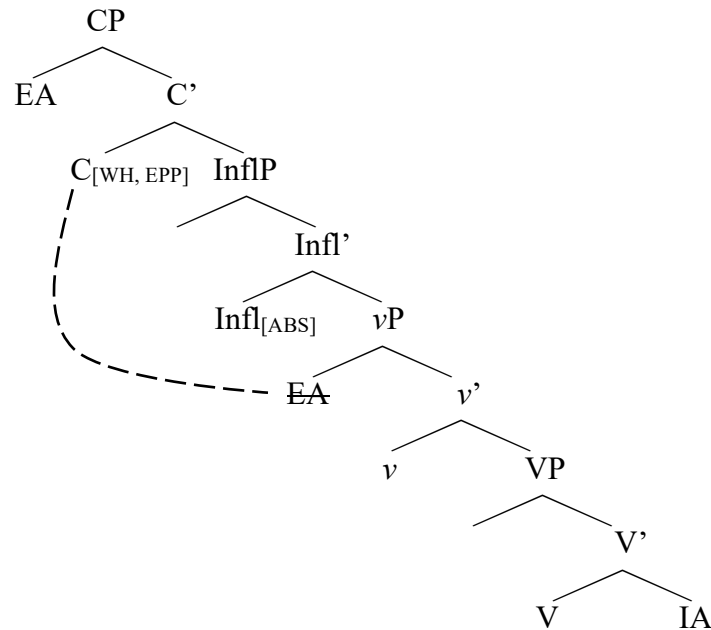
Remember once again that AF can *only* be used when the subject is extracted—any other configuration prohibits the use of AF. This is a crucial difference between AF and other voices in Kaqchikel (like active and passive): AF is available only in a strictly defined set of environments, whose complement set is the 'elsewhere' set. If the EA of a transitive clause bears an A'-feature, AF *must* be used. Partly because of this Last Resort flavor, I will analyze AF as the result of the application of the operation Exfoliation, which Pesetsky (2021) proposes is itself a Last Resort operation.

   Let us briefly examine the proposal underpinning a derivational operation of structure removal like Exfoliation. Pesetsky 2021 notes that researchers for the past four decades of generative syntax have assumed the proposals in Kiparsky & Kiparsky (1971) (henceforth

K&K) and Bresnan (1972) regarding clause size. These works proposed, in the context of the Standard Theory (Chomsky 1965), that different predicates select for complements of different sizes; i.e., clause versus NP. Their arguments were based on requirements of the theory at the time, in particular the need to have different clause types differentiated at the level of Deep Structure, which was the input to semantic interpretation (for the hypothesis that semantics interprets Deep Structure, not Surface Structure, see Katz & Postal 1964).[36] Pesetsky contends that the main arguments forming the backbone of K&K and Bresnan are no longer tenable under modern Minimalist approaches, so the consequences of reviving a derivational approach to clause size à la Lees (1963) and Rosenbaum (1965, 1967) deserve to be explored.[37]

More specifically, Pesetsky proposes that clauses can be reduced as a response to locality problems arising in the course of the derivation (see also Müller 2015). The

---

[36] Note that K&K, for instance, do not reject structure removal entirely (see K&K 159-163). Rather, they discuss the difference between factive and non-factive complement clauses and maintain that the contrast between them results from a Deep Structure difference. In other words, on this proposal factive and non-factive complement clauses specifically are not derivationally related.

[37] An operation like Exfoliation traces its roots to proposals by Haj Ross as well, as noted by Pesetsky (2017). In chapter 3 of Ross 1967, a tree pruning transformation is proposed that deletes tree branches that do not dominate anything, whereas Ross 1969b discusses "Node Deletion", a transformation that deletes tree nodes in order to feed or bleed the application of other transformations. Exfoliation is similar in spirit to Node Deletion. Contemporaneously to Pesetsky, Müller (2015) proposes a local operation Remove which is the mirror image of Merge and delivers similar results to Exfoliation, although the precise technical implementation differs. Furthermore, while Pesetsky focuses mostly on structure removal at the CP level, Müller concentrates mostly on the Voice/*v* level—in a sense, then, the present proposal applies Pesetsky's structure removal operation (Exfoliation) to the empirical domain explored by Müller.

particular structural description which triggers the application of Exfoliation is shown in (96). Exfoliation occurs when a goal α is inaccessible to movement resulting from probing by a probe β outside of the phasal domain in which α is located.

(96)   *Exfoliation* (adapted from Pesetsky 2021: 12)[38]

    a.   **Structural description:** β … [YP(PHASE) …. [γP(NON-PHASE) … α]], where
        (i) YP is the phase that dominates α, but not β,
        (ii) α occupies the edge of γP, and
        (iii) a movement-triggering probe on β has located α as its goal

    b.   **Structural Change:** Replace YP with γP

I also adopt the conception of phasehood penetrability and impenetrability in Pesetsky's work. Note that this conception of phasehood departs from the assumption that the complement of a phase head is unavailable for all syntactic operations (compare this with the approaches to phase impenetrability discussed above (93)):

(97)   *Probing across a clause boundary; CP* (Pesetsky 2021: 12)

    a.   **Phase penetrability**: A probe π with an EPP property can locate a goal γ across a CP boundary, even if γ does not occupy the edge of that CP …

    b.   **Phase impenetrability**: … but γ can move to π only if it occupies the edge of its clause.

---

[38] One element of Pesetsky (2021)'s definition of Exfoliation that I do *not* adopt is γP becoming the phase head following Exfoliation. Pesetsky's Structural Change is, to quote, "Replace YP with *γ* P*, which takes the phasal property of its predecessor*". This part of the definition was not specified in earlier versions of the proposal (Pesetsky 2017) and is used to account for why the head of the *to*P phrase is only spelled out when Exfoliation has applied (see Pesetsky 2021: 13's *Exposure*). It seems to me that removal of the phase *in toto* via Exfoliation is simpler than formalizing how γP would become a phase after Exfoliation, when it previously was not a phase—though, of course, we would need an alternative account for the conditions in which the head of *to*P is pronounced. We will observe that for our purposes, it is important that γP not become a phase head, given the omnivorous nature of the Agree probe on Infl in the Agent Focus construction (more on this in 2.4).

In brief, a probe can locate a goal, but movement of the goal is prohibited unless it occupies the phase edge.

I will make one additional assumption regarding phase impenetrability: feature sharing mediated by an Agree relation cannot proceed between a probe and a goal that is inside of the complement of a phase head:

(98)   *Probing across a clause boundary—no feature-sharing established*
  a. **Phase penetrability**: A probe π can locate a goal γ across a phase boundary, even if γ does not occupy the edge of that phase …
  b. **Phase impenetrability**: … but a feature sharing relation cannot be established between π and γ.

As we will see, I will propose that Infl^0 can only establish a feature sharing relation with the internal argument if the VoiceP domain has been Exfoliated. This will derive the fact that probing by Infl^0 ends up being omnivorous in the Agent Focus configuration, but not in regular transitive clauses (see below and section 2.4).

Returning to Exfoliation and Pesetsky's proposal, one of its main explananda is the syntactic behavior of Raising-to-Object/ECM (RtO) and Raising-to-Subject (RtS). In RtO and RtS, a probe in a matrix clause (v^0 for RtS; V^0 in RtO) probes for a goal in the embedded clause, across the embedded C^0 phase. Exfoliation of the embedded CP allows for the goal to move to the specifier of the relevant probe. Pesetsky's Exfoliation account can explain some surprising properties of these constructions. Let us discuss RtS briefly.

Pesetsky notes that the distribution of nominal subjects of infinitivals has been important for defending the existence of the Case Filter: Raising is triggered by the nominals' need to be Case licensed. However, non-nominals such as clauses (99)c-d, verbal predicates (99)e-f, and locatives (99)g-h share the same distribution with nominals in this configuration. A sample of the contrasts discussed in Pesetsky 2019 is given below:

(99)   *Subjects of infinitivals in English—no Case Filter explanation available*

      a.   Mary seems to speak French well.

      b.   *It was believed Mary to speak French well.

      c.   [That the world is round] seems to be a tragedy.

      d.   *It was believed [that the world is round] to be a tragedy.

      e.   [Even more important than linguistics] seems to be the fate of the planet.

      f.   *It was believed [even more important than linguistics] to be the fate of the planet.

      g.   [In this room] seems to have been found the finest examples of Athenian sculpture.

      h.   *It was believed [in this room] to have been found the finest examples of Athenian sculpture.

Pesetsky argues that the deviancy of the starred examples above cannot be the result of the Case Filter—it is unclear how all the subjects in the starred examples would fall under its purview. In contrast, Pesetsky attributes the ungrammatical status of the starred examples above to *illegal applications of Exfoliation*.

Let us unpack this slightly. Crucial for Pesetsky's approach is the idea that infinitivization can *only* be achieved via Exfoliation. In other words, his theory is highly derivational, and every clause starts out as a full (finite) clause. Consider Pesetsky's analysis, which is sketched below. The embedded subject occupies the specifier position of *to*P. A higher probe β (for instance, a v^0 probe in RtS) finds the embedded subject, but that embedded subject is unable to raise. Exfoliation applies (rendering the embedded clause infinitival as a side-effect) and the embedded subject is thus able to raise:[39]

(100) *Infinitivization achieved via Exfoliation*

    a. [… β$_{[F]}$ … [CP$_{[phase]}$ [TP [toP Subj$_{[F]}$ …]]]]   *Subj cannot raise*

    b. [… β$_{[F]}$ … [toP Subj$_{[F]}$ …]].   *Exfoliation; CP-TP deleted; infinitivization*

    c. [βP Subj [β … [toP <Subj>]]]   *Subj raises from infinitival clause*

---

[39] The empirical consequences of placing the subject in such a position (as opposed to Spec,TP) are detailed in Pesetsky's work, but going over them here would take us too far afield.

The starred examples, however, do not meet the structural description of Exfoliation, since there is no probe in the matrix clause whose presence would trigger raising of the embedded subject. In other words, the starred examples are illicit because they cannot legitimately surface as infinitival: Pesetsky proposes that passive verbs do not bear a probe seeking to attract the embedded subject, so the application of Exfoliation is not triggered (see Pesetsky 2021: example (35)).

While still a novel proposal, Exfoliation has the potential of providing a unified solution to the puzzle of the parallel behavior of all the subject types under RtS (as well as having consequences in other empirical domains, e.g., Comp-trace effects). I will not provide any additional arguments for the availability of Exfoliation as an operation in the narrow syntax, since the discussion would take us too far afield. I point the reader instead to Pesetsky 2021 for extensive argumentation.

At this juncture, note that the definition of Exfoliation in (96) does not restrict itself to a particular identity for the probe and goal. Whereas Pesetsky discusses applying Exfoliation only to the C^0 phase, the proposal advanced here extends the application of Exfoliation to the verbal domain ($Voice^0$ in Kaqchikel), an extension that is predicted to exist. In other words, given the Exfoliation proposal, if a language's particular grammar is such that an accessibility problem arises at the $Voice^0$ level, then just like at the CP level, Exfoliation is predicted to apply, as a Last Resort. For Kaqchikel specifically, an EA bearing an A'-feature cannot move out of the verbal domain, since the IA occupies the sole escape hatch on $Voice^0$. As a result, Exfoliation applies:[40]

---

[40] I discuss the fate of Spec,VoiceP in section 2.5.

(101) *Probing across a clause boundary; VoiceP*

    a.   **Phase penetrability**: A probe π with an EPP property can locate a goal γ across a VoiceP boundary, even if γ does not occupy the edge of that VoiceP

        …

    b.   **Phase impenetrability**: … but γ can move to π only if it occupies the edge of its domain.

(102) *Exfoliation of Voice in Kaqchikel = AF*

    a.   **Structural description**: C … [VoiceP$_{(PHASE)}$ …. [vP$_{(NON-PHASE)}$ … EA]], where

        (i) VoiceP is the phase that dominates EA, but not C,

        (ii) EA occupies the edge of vP, and

        (iii) a movement triggering probe on C has located EA as its goal

    b.   **Structural Change**: Replace VoiceP with vP

    c.   *Structure without exfoliation*

93

d.   *Exfoliation applies*

CP

C[WH]   InflP

Infl'

Infl   *v*P

EA[WH]   VP

V   IA

(*now accessible, given
the removal of phase
boundary*)
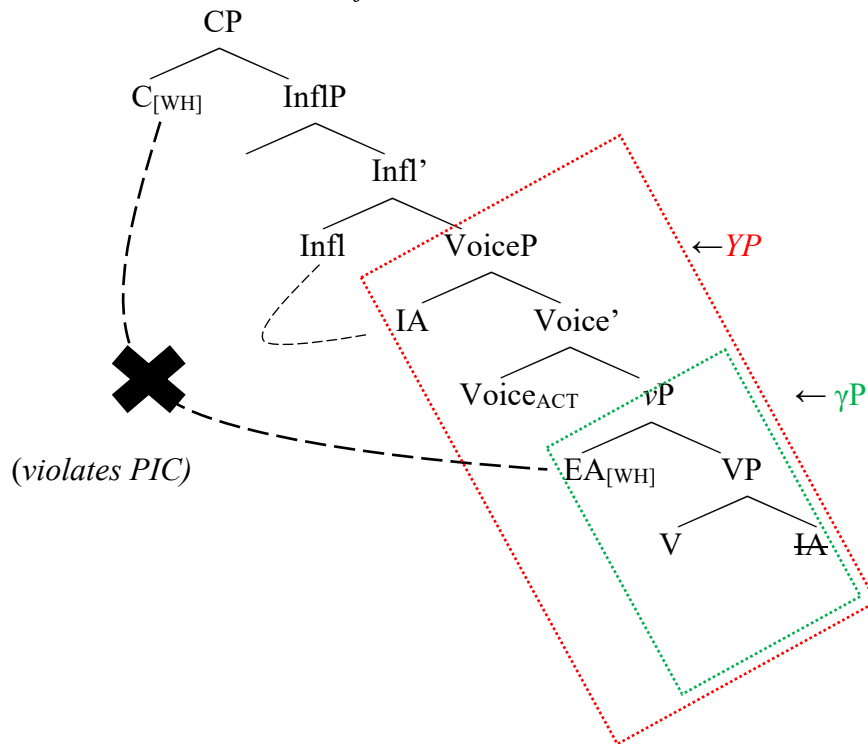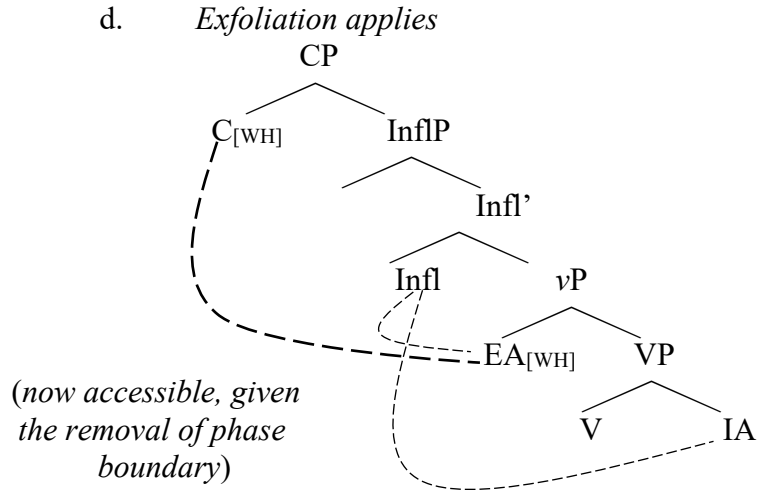
It is worth pointing out again that (102)a-b do not represent a set of stipulations that are specific to Kaqchikel and the data that have been central to our discussion. Instead, (102)a-b are an implementation of Pesetsky's proposal *as is* to the Kaqchikel VoiceP phase (modulo the provisos about the status of *v*P as a phase post-application of Exfoliation; see footnote 38).

The application of Exfoliation as a Last Resort operation has several consequences for Kaqchikel:

(103)   *Consequences of Exfoliation*
   a.   The Voice^0 phase is removed, allowing for the EA to move after being targeted by the C[WH] probe.
   b.   The EA and IA become phasemates and can both be targeted by Infl^0.
   c.   Voice_ACT—the head/landing-site of the clitic doubling chain <Voice_ACT, EA> that results in ergative (Set A) agreement—is deleted.
   d.   The higher copy of the IA is deleted. Thus, the clitic doubling chain <Infl^0, IA> is disrupted.

Note that the application of Exfoliation disrupts both clitic doubling chains that would have resulted in agreement. Set A agreement (the result of clitic doubling between Voice^0 and the EA), as well as set B agreement (the result of clitic doubling between Infl^0 and the IA) are both disrupted as a result of Exfoliation (99)c-d.

94

A question that arises is how the EA and IA can be targeted by $Infl^0$ following Exfoliation, since they had already been the target of $Infl^0$ and $Voice_{ACT}$ respectively. Recall that we are following a suggestion in Preminger 2014's analysis that Set A (as well as Set B) agreement are doubled clitics on the probing head. It has been observed that doubled clitics behave like the heads of A-chains insofar as they render the lower copy (the clitic-doubled DP) a non-intervener (Anagnostopoulou 2003). Doubled clitics themselves are unlike heads of (phrasal) A-chains in that they are not themselves interveners or candidates for A-probing (i.e., viable goals for Agree)—see Preminger 2009. I assume that Exfoliation deleting portions of the clitic doubling chain renders the arguments once again available as targets for Agree, since they are no longer part of any clitic doubling chain. This goes some way towards explaining why the EA can be a target for $Infl^0$ in the AF clause even though it cannot be targeted for such agreement in a regular active transitive clause. I will return in section 2.4 to the consequences of analyzing AF as Exfoliation in more detail.

Most importantly for our present purposes, we can now begin to understand why Active-AF, AF-Active, and Passive-AF mismatches would be allowed under ellipsis. These configurations comply with the syntactic identity condition proposed in (89). An AF antecedent or sluice instantiates a structure in which the Voice layer has been removed completely. When an AF antecedent or sluice is paired with $Voice_{ACT}$ or $Voice_{PASS}$ in an ellipsis configuration, there is no featural clash which would lead to a violation of (89), since there is nothing left in the AF clause for these heads to clash with. The examples involving Active-AF, AF-Active, and Passive-AF mismatches in sluicing are as shown below, then: there is no clash in $Voice^0$ heads.

(104) *Voice mismatch under sluicing—featural non-distinctness allowed*

Antecedent	Sluice
a. Voice[F: ACT]	Ø (AF)
b. Ø (AF)	Voice [F: ACT]
c. Voice [F: PASS]	Ø (AF)

We have thus arrived at an explanation for why some voice mismatches in Kaqchikel are allowable, whereas other combinations of voice mismatches are banned (e.g., all English voice mismatches and AP-Active mismatches in Kaqchikel). We will zoom into each of the mismatches in detail in 2.3.4 below.

Before moving on, an alternative to the Exfoliation account of Agent Focus is worth considering, since it might deliver similar empirical coverage when it comes to the Kaqchikel sluicing examples. Instead of the analysis proposed above—that a specific structural configuration leads to the removal of VoiceP via Exfoliation—could we not assume that a truncated clause can be constructed when the argument merged in Spec,vP bears an A'-feature? In other words, instead of VoiceP merging with Infl^0 in such a case, vP would merge with Infl^0. An Agent Focus clause would thus be this extraordinary derivation lacking the VoiceP layer from the start—a derivation that is ill-formed otherwise.

Empirically, this alternative might be on equal footing with the Exfoliation approach, since non-distinctness would be satisfied in the mismatch cases that were analyzed. However, one might worry that this alternative account misses the generalization that the extraction restriction—the fact that the EA in an active transitive cannot move from the Voice domain—seems to *trigger* a specific Last Resort process in order for the restriction to be circumvented. In other words, this alternative analysis would miss the apparent cause-effect relation between the extraction restriction and the use of Agent Focus (in our terms,

the application of Exfoliation). In order to maintain the cause-effect relation under this alternative, one would need to posit that a derivation that fails due to the extraction restriction triggers a *new derivation* where VoiceP is not merged as a Last Resort. At this juncture in comparing the two approaches, a debate could be had on whether one proposal is more conceptually palatable than the other, but such a discussion would take us too far afield. I leave for the future, then, further consideration of this alternative.[41]

To summarize this subsection, I laid out my assumptions regarding (i) the clausal structure of transitives in Kaqchikel and (ii) the source of the extraction restriction on A'-movement of transitive subjects that gives rise to AF. I showed that AF can be derived as the clausal structure that results from the removal of VoiceP via the operation Exfoliation. Before delving deeper into the voice mismatches that concern us in this chapter, let us take a brief excursus into how the proposed analysis of AF as Exfoliation derives the morphemes that appear in Kaqchikel verbal stems.

2.3.3 *Voice morphology*

Let us now consider the appearance of the AF morpheme and other exponents of voice and transitivity in the Kaqchikel verbal stem. As we noted before, Patzún Kaqchikel verbal templates have a final slot that I have glossed as the specification of voice or transitivity:[42]

---

[41] Thank you to Omer Preminger and Masha Polinsky for raising this alternative.

[42] I depart from available descriptions of Kaqchikel morphology (García Matzar 2007) in assuming that the vowels in the -Vj and -V̈x suffixes are segments in the suffix, instead of being segments in the stem of affixation; e.g., García Matzar 2007: 71 gives the relevant passive as -x.

(105) *Voice morphology in Kaqchikel*

Exponence	Verb Type	Voice
-Ø	root CVC	active
-Vj	derived	active
root V̈ → V	root CVC	passive
- V̈x	derived	passive
-Vn	all types	AP
-o/u	root CVC	AF/AP$_{OBL}$
-Vn	derived	AF/AP$_{OBL}$

First, I assume that verb initial order in Kaqchikel is derived via head-movement of V^0 to Infl^0 (see Clemens & Coon 2018, England 1991, Douglas et al. 2017 for word order across Mayan). Second, I assume that lexical insertion is a post-syntactic process, as in Distributed Morphology approaches (Halle & Marantz 1993, Embick 2015). I propose that the suffixes we observe in the Kaqchikel verbal stem instantiate Vocabulary Insertion (VI) into different flavors of v^0. In contrast, there is no morpheme that instantiates insertion into Voice^0.

In discussions of Kaqchikel morphology (see García Matzar & Rodríguez Guaján 1997, Patal Majzul et al. 2000, García Matzar 2007), the appearance of one or another allomorph of the AF/AP$_{OBL}$ and the simple passive is dependent on properties of the stem of affixation. Descriptively, the relevant distinction is between so-called "root/underived/radical" transitives and "non-root/derived" verbal stems. For the Patzún speakers I have worked with, root/underived/radical transitives take a -o/-u morpheme, whereas non-root/derived transitives—i.e., those transitive stems that surface with a -Vj segment in the active voice— take -Vn.[43] Similarly for the exponence of the passive, the vowel in root/underived/radical

---

[43] The specific terminology might vary, but the basic distinction between two types of verbal stems is used to describe morpheme alternations that are found across the entire Mayan family; see, for instance, Polian 2017.

transitive roots undergoes vowel tensing, while non-root/derived transitives combine with a -Ṽx suffix. I propose that the oft-invoked distinction between root/underived/radical and non-root/derived stems amounts to the existence of two different verbal root classes. In other words, there is a list of verbal roots that combine with a specific transitive v^0, whereas all other verbal stems combine with a different transitive v^0. The morphological differences we observe follow from different VI rules for different v^0 heads. Let us observe the derivations I am proposing explicitly.

First consider root/underived/radical transitives such as *löq'* 'buy', *tz'ët* 'see', *tïk* 'to plant'. In the active voice, these verbs surface without any suffix. The proposed structure for each example appears on the right; I propose that the verb roots in this class combine with a transitive v^0 that I label v_{trans1}:

(106) *So-called "root/underived" transitives*

    a. *Active voice*
      x-Ø-u-löq'. $[\text{Voice}_{\text{ACT}}\ [v_{\text{trans1}}\ [\text{V}_1]]]$
      COM-B3S-A3S-buy

    b. *Passive voice*
      x-Ø-loq' $[\text{Voice}_{\text{PASS}}\ [v_{\text{trans1}}\ [\text{V}_1]]]$
      COM-B3S-buy.PASS

    c. *AP voice*
      x-Ø-loq'-on $[\text{Voice}_{\text{AP}}\ [v_{\text{trans1}}\ [\text{V}_1]]]$
      COM-B3S-buy-AP

    d. *AF/AP_{OBL} voice*
      x-Ø-loq'-o $[v_{\text{trans1}}\ [\text{V}_1]]$ *after Exfoliation*
      COM-B3S-buy-AF/APO

I assume that there exist strict conditions on the locality of allomorphy, such that only elements immediately adjacent to the head undergoing VI can condition the selection of one allomorph over another in the post-syntactic module. In other words, the insertion of a particular morpheme in v^0 is conditioned by its immediate environment, created via head

movement.[44] The following Vocabulary Insertion rules derive the attested exponence for different voices in this type of Kaqchikel data:

(107) *Vocabulary Insertion rules in Kaqchikel; so-called "root/underived" transitives*

a.	v_{trans1}	↔	-Ø	/	_Voice$_{ACT}$	(106)a
b.	v_{trans1}	↔	[+tense]	/	_Voice$_{PASS}$	(106)b
c.	v_{trans1}	↔	-Vn	/	_Voice$_{AP}$	(106)c
d.	v_{trans1}	↔	-o			(106)d

Note that the insertion of the AF/AP$_{OBL}$ is the elsewhere case, inserted when v_{trans1} is not in the context of any flavor of Voice^0.

Consider now a second set of data, involving verbal roots like *k'ay* 'sell', *q'et* 'hug', *ray* 'desire', etc. These roots surfaces with a -Vj suffix in the active voice; i.e., *k'ayij, q'etej*, and *rayij* respectively. I propose that the verb roots in this class combine with a different transitive v^0 than those in (106), labelled v_{trans2} in the structures below:

(108) *So-called "derived" transitives*

    a. *Active voice*

      x-Ø-u-k'ay-ij.   [Voice$_{ACT}$ [v_{trans2} [V$_2$]]]

      COM-B3S-A3S-sell-ACT

    b. *Passive voice*

      x-Ø-k'ay-ïx   [Voice$_{PASS}$ [v_{trans2} [V$_2$]]]

      COM-B3S-sold-PAS

    c. *AP voice*

      x-Ø-k'ay-in   [Voice$_{AP}$ [v_{trans2} [V$_2$]]]

      COM-B3S-sell-AP

    d. *AF/AP$_{OBL}$ voice*

      x-Ø-k'ay-in   [v_{trans2} [V$_2$]]   *after Exfoliation*

      COM-B3S-sell-AF/APO

---

[44] I use V^0 instead of a √ROOT here for ease of exposition (see Harley 2014 for discussion on the status of √ROOTs in the grammar and chapter 4 for my proposal on the portion of the identity condition that is specific to them).

In a manner parallel to the insertion rules for the first set of data in (107), I propose that the exponent we observe here for voice/transitivity is the insertion of transitive v^0 (here v_{trans2}) in different structural contexts:

(109) *Vocabulary Insertion rules in Kaqchikel; so-called "derived" transitives*

a.	v_{trans2}	↔	-Vj	/ ___Voice$_{\text{ACT}}$	(108)a
b.	v_{trans2}	↔	-V̈x	/ ___Voice$_{\text{PASS}}$	(108)b
c.	v_{trans2}	↔	-Vn		(108)c
d.	v_{trans2}	↔	-Vn		(108)d

Once again, the insertion of the AF/AP$_{\text{OBL}}$ is the elsewhere case. In this case, the elsewhere morpheme is inserted when v_{trans2} is not in the context of Voice^0; i.e., when Voice^0 has been removed via Exfoliation.

Finally, consider transitive stems that involve causativization of an intransitive root, e.g., *kamsaj* 'to kill', which is composed of the intransitive root *kam* 'to die' and two additional suffixes *-sa* and *-Vj*. I assume that when these suffixes are stacked, the illicit VV sequence that would result is reduced in the phonological component:

(110) *Causative transitives derived from intransitive stems*

    a. *Active voice*
      x-Ø-u-kam-sa-j.   [Voice$_{\text{ACT}}$ [v_{trans2} [v_{intrans} [V]]]]
      COM-B3S-A3S-die-CAUS-ACT

    b. *Passive voice*
      x-Ø-kam-is-äx   [Voice$_{\text{PASS}}$ [v_{trans2} [v_{intrans} [V]]]]
      COM-B3S-die-CAUS-PAS

    c. *AP voice*
      x-Ø-kam-sa-n   [Voice$_{\text{AP}}$ [v_{trans2} [v_{intrans} [V]]]]
      COM-B3S-die-CAUS-AP

    d. *AF/AP$_{OBL}$ voice*
      x-Ø-kam-sa-n   [v_{trans2} [v_{intrans} [V]]]   *after Exfoliation*
      COM-B3S-die-CAUS-AF/APO

I propose that the *-sa* morpheme in examples like those shown above is the exponent of v_{intrans} in the context of the transitivizing v_{trans2}:

(111) *Vocabulary Insertion rules in Kaqchikel; causatives derived from intransitive stems*

a.	$v_{intrans}$	$\leftrightarrow$	-sa	/	$\_ v_{trans2}$	(110)a
	v_{trans2}	$\leftrightarrow$	-Vj	/	$\_ \text{Voice}_{ACT}$	
b.	$v_{intrans}$	$\leftrightarrow$	-(i)sa	/	$\_ v_{trans2}$	(110)b
	v_{trans2}	$\leftrightarrow$	-V̈x	/	$\_ \text{Voice}_{PASS}$	
c.	$v_{intrans}$	$\leftrightarrow$	-sa	/	$\_ v_{trans2}$	(110)c
	v_{trans2}	$\leftrightarrow$	-Vn			
d.	$v_{intrans}$	$\leftrightarrow$	-sa	/	$\_ v_{trans2}$	(110)d
	v_{trans2}	$\leftrightarrow$	-Vn			

To summarize, the VI rules in (107), (109), and (111) account for the different voice/transitivity morphemes that appear across different verbal stem types in Kaqchikel.

We return to sluicing in the following subsection, where I will show how the analysis of AF defended in this chapter gives us a handle on deriving the attested voice mismatches in Kaqchikel.

2.3.4. Deriving voice mismatches in Kaqchikel sluicing

I will now lay out some background assumptions regarding the syntactic analysis of sluicing in general. I assume a PF deletion approach to ellipsis (Ross 1969a, Merchant 2001; see van Craenenbroeck and Merchant 2013, Lasnik and Funakoshi 2019, Saab 2021 for discussion).[45] I also assume that an [E]-feature on specific heads licenses ellipsis (see chapter 1; Merchant 2001) and that *wh*-remnants escape PF-deletion by moving out of the ellipsis site. Recall that case matching effects and the P-stranding generalization have been taken as evidence for this movement step (for an *in-situ* approach to sluicing, see Abe

---

[45] I use the term PF-deletion since it is familiar from recent and contemporary literature. In Distributed Morphology approaches, some have taken PF-deletion to mean suppression of Vocabulary Insertion (Saab 2021) or application of a morphological operation like Obliteration on all elements inside the ellipsis site (Murphy 2016); for an important precursor to these approaches, see Wasow 1972's *Empty Structure Hypothesis*. These distinctions are not crucial for our present discussion; see chapter 1.

2015). There is a Kaqchikel-internal diagnostic that leads us to the same conclusion, which

I now discuss.

Recall that certain oblique elements are introduced by a preposition-like element in

Kaqchikel called a relational noun (RN). With this in mind, note that *wh*-movement of

adjuncts involving a RN+DP structure obligatorily triggers pied-piping with inversion

(henceforth PPI; see Aissen 1996 for Tsotsil). A schematic of this phenomenon is shown

below:

(112) *Obligatory pied-piping with inversion in Kaqchikel*
    a.   V O S [RN + DP$_{[WH]}$]   *structure prior to movement*
    b.   [DP$_{[WH]}$ +RN]$_1$ V O S t$_1$ *wh-movement*

The baseline data showing PPI are given below. Note that it is impossible to *wh*-extract

these adjuncts without PPI. I illustrate with a benefactive, but this phenomenon holds with

other adjunct types as well:

(113) *Obligatory pied-piping with inversion in Kaqchikel simple wh-questions*
    a. Maria x-Ø-u-löq'   jun machït chin a   Pedro.
       Maria COM-B3S-A3S-buy a   machete RN   CLF Pedro
       'Maria bought a machete for Pedro.'   *baseline*

    b. Achoj chin x-Ø-u-löq'   jun machït   xta Maria?
       WH   RN COM-B3S-A3S-buy a   machete CLF Maria
       'Who did Maria buy the machete for?'   *wh-movement and PPI*

    c. *Chin achoj x-Ø-u-löq'   jun machït xta Maria?
        RN   WH   COM-B3S-A3S-buy a   machete CLF Maria
       'Who did Maria buy the machete for?'   **wh-movement and no PPI*

What is crucial for our purposes is that under sluicing, the same obligatory PPI is observed.

Consider the relevant data below:

(114) *Obligatory pied-piping with inversion in Kaqchikel sluicing*
　　　a. Maria x-Ø-u-löq'　　　jun machït.　Man w-etama-n　　ta　[achoj chin]₁
　　　　 Maria COM-B3S-A3S-buy a 　machete NEG A1S-know-PERF NEG 　WH 　RN
　　　　 <… t₁>.

　　　　 'Maria bought a machete. I don't know who for.'
　　　b. *Maria x-Ø-u-löq'　　　jun machït.　Man w-etama-n　　ta　[chin achoj]₁
　　　　 Maria COM-B3S-A3S-buy a 　machete NEG A1S-know-PERF NEG 　RN 　WH
　　　　 <… t₁>.

　　　　 Intended: 'Maria bought a machete. I don't know who for.'

In a nutshell, we can straightforwardly account for the inverted nature of the RN+DP[WH] in examples like those above if we assume that the *wh*-phrase has moved. In contrast, a base-generation or WYSIWYG approach would have to stipulate inversion separately for sluicing and for non-elliptical cases.

　　Having established our assumptions regarding the analysis of sluicing proper, let us now illustrate a full derivation of the Voice mismatches under sluicing that we have discussed so far, starting with the Active-AF mismatch.

　　The structure of antecedent and target clause is seen below at each step of the derivation. In the target clause, I only focus on the relevant chunk of structure (the ellipsis site is greyed out):

(115) *Active-AF mismatch*
a.　　　*Prior to Exfoliation in the target clause*
[CP [C [InflP [Infl[ABS] [VoiceP IA [Voice_ACT [ERG, EPP] [vP EA [v [VP [V IA]]]]]]]]]]　*antecedent*
[CP [C[WH, EPP, E] [InflP [Infl[ABS] [VoiceP IA [Voice_ACT [ERG, EPP] [vP EA[WH] [v [VP [V IA]]]]]]]]]]]
　　　　　　　　　　　　　　　　　　　　　　　　　　　　　target clause
b.　　　*After Exfoliation in the target clause*
[CP [C [InflP [Infl[ABS] [VoiceP IA [Voice_ACT [ERG, EPP] [vP EA [v [VP [V IA]]]]]]]]]]　*antecedent*
[CP [C[WH, EPP, E] [InflP [Infl [vP EA[WH] [v [VP [V IA]]]]]]]]]　　　　*target clause*
c.　　　*Sluicing*
[CP [C [InflP [Infl[ABS] [VoiceP IA [Voice_ACT [ERG, EPP] [vP EA [v [VP [V IA]]]]]]]]]]　*antecedent*
[CP EA[WH] [C[WH, EPP, E] [InflP [Infl [vP EA[WH] [v [VP [V IA]]]]]]]]]　　*target clause*

Sluicing is acceptable in (115)c above because the elided clause is featurally non-distinct from the antecedent clause, complying with (89). We therefore observe that an active clause can mismatch with AF.

There is, however, an important alternative that we should consider at this juncture. Suppose that apparent Active-AF mismatches are actually instances of island repair by ellipsis:

(116)   *Active-AF mismatch (putative alternative analysis)*
   In apparent Active-AF mismatches, the sluice is actually in the *active voice*. The derivation converges because extraction of an EA *wh*–phrase from an active clause is an island violation. Crucially, island violations are repaired under sluicing.[46]

The analysis in (116) would take the acceptability of the examples that I have been referring to as Active-AF mismatches to instantiate an Active-Active match plus island repair by ellipsis. It has long been observed that island violations seem to be repaired through sluicing (Ross 1969a, Lasnik 2001, Merchant 2001, Merchant 2008, Lasnik 2009).

(117)   *Island repair under sluicing*
   a. *I believe the claim that she saw someone, but they don't know **who$_1$** I believe the claim that she saw **t$_1$**.
   b. I believe the claim that she saw someone, but they don't know **who$_1$** <~~I believe the claim that she saw~~ **t$_1$**>.
   c. *Tayce kissed a guy who bit one of my friends, but Lawrence doesn't know [**which one of my friends**]$_1$ Tayce kissed a guy who bit **t$_1$**.
   d. Tayce kissed a guy who bit one of my friends, but Lawrence doesn't know [**which one of my friends**]$_1$ <~~Tayce kissed a guy who bit~~ **t$_1$** >.

---

[46] Exactly *how* sluicing repairs island violations has been a topic of ample discussion (see van Craenenbroeck and Merchant 2013, Lasnik and Funakoshi 2019, Mendes 2020). I present one possible mechanism to encode island repair in order to illustrate the analysis for Kaqchikel voice mismatches. For arguments for an opposing view—that there exists no island repair by ellipsis—see Barros et al. 2014 and Abels 2019a for discussion. Naturally, if this opposing view is correct, then the putative alternative analysis I am evaluating here can be discarded outright.

e. *Ellie bought beans and corn at the market, but I don't know [**which beans**]₁ she bought **t₁** and corn at the market.

f. Ellie bought beans and corn at the market, but I don't know [**which beans**]₁ <~~she bought t₁ and corn at the market~~.>

One could dismiss this line of argumentation immediately if Kaqchikel sluicing did not repair island violations generally. However, Kaqchikel behaves as expected (see Mendes & Ranero 2021). Observe first a coordinate island example below:

(118) *Coordinate island structure*

   a. Ana x-Ø-u-löq'   ch'op   chqa' saq'ul.
      Ana COM-B3S-A3S-buy pineapple and   bananas
      'Ana bought pineapples and bananas.'

   b. *Achike₁ x-Ø-u-löq'   ch'op   chqa' t₁?
       what   COM-B3S-A3S-buy pineapple and
      Intended: 'What is the thing such that she bought pineapples and that thing?'

   c. *Achike₁ x-Ø-u-löq'   t₁ chqa' saq'ul?
       what   COM-B3S-A3S-buy   and   banana
      Intended: 'What is the thing such that she bought that thing and bananas?'

(119) *Island violation repaired under sluicing; coordinate island*

   Ana x-Ø-u-löq'   ch'op   chqa' saq'ul,   po man w-etama-n   ta
   Ana COM-B3S-A3S-buy pineapple and   bananas but NEG A1S-know-PERF NEG
   achike saq'ul!
   what   banana
    'Ana bought pineapples and bananas, but I don't know which bananas!'

Next consider an island violation involving an adjunct *if*-clause. Sluicing can repair this type of island violation as well:

(120) *If-clause island structure*

   a. Yïn n-Ø-in-tïj   ri   nu-xajab' si ma Juan n-Ø-u-k'ay-ij
      1S INC-B3S-A1S-eat DET A1S-sandal if CLF Juan INC-B3S-A3S-sell-ACT
      ri   ru-ch'ich'.
      DET A3S-car
      'I will eat my sandal if Juan sells his car.'

   b. Achike₁ n-Ø-in-tïj   t₁ si ma Juan n-Ø-u-k'ay-ij   ri   ru-ch'ich'?
      what   INC-B3S-A1S-eat   if CLF Juan INC-B3S-A3S-sell-ACT DET A3S-car
      'What will I eat if Juan sells his car?'

    c. *Achike n-Ø-in-tïj   nu-xajab' si ma Juan n-Ø-u-k'ay-ij   t₁?

       what   INC-B3S-A1S-eat A1S-sandal if CLF Juan INC-B3S-A3S-sell-ACT

       Intended: 'What is the thing such that I would eat my sandal if Juan sold that
              thing?'

(121) *Island violation repair under sluicing; if-clause island*

    a. Yïn n-Ø-in-tïj   ri nu-xajab' si ma Juan k'o

       1S INC-B3S-A1S-eat DET A1S-sandal if CLF Juan EXIST

       n-Ø-u-k'ay-ij=el.   Aw-etama-n   achike?

       EXIST INC-B3S-A3S-sell-ACT=DIR A2S-know-PERF what

       'I will eat my sandal if Juan sells something. Do you know what?'

Since island violations are indeed repaired under sluicing in Kaqchikel, we must seriously

consider the putative alternative analysis in (116). Under that approach, the extraction

restriction arises because extracting the EA is an island violation: Movement of the EA out

of the phase is illicit, because the IA occupies the single specifier of the phase. Let us

assume that this kind of illicit movement is recorded in the derivation via a diacritic within

the phasal domain. I mark this diacritic below as an asterisk (Chomsky 1972; see Lakoff

1972, Lasnik 2001, Merchant 2008):[47]

(122) *Extraction Restriction*

    [CP EA[WH] [C[WH, ~~EPP~~] [InflP [Infl[ABS] [* VoiceP IA [VoiceACT [ERG, ~~EPP~~] [vP ~~EA~~ [WH] [v
    [VP [V ~~IA~~]]]]]]]]]]]

Under this approach, sluicing would remove the InflP clause, leading to the deletion of the

* diacritic on VoiceP as well. If this is the case, whenever we have assumed that extraction

---

[47] An alternative to the diacritic mechanism is the idea that movement out of an island creates a PF problem.

Therefore, ellipsis at PF gets rid of the externalized material that would have given rise to the violation (see

for example Hornstein et al. 2007 and Mendes 2020 for discussion). If this encoding of island violations is

on the right track, we would need to assume that the extraction restriction in Kaqchikel arises because of a

PF issue, as opposed to a derivational issue in the syntax.

of an EA necessarily leads to the use of AF in the sluice, we have been incorrect. These examples could instead involve extracting the EA from an *active* clause, leading to an island violation. Since the example involves a sluice, however, the extraction would ultimately be allowed. Therefore, there is no mismatch at all—both clauses are in the active voice:

(123) *Active-AF mismatches are Active-Active and involve island repair by sluicing*

[CP [C [InflP [Infl[ABS] [VoiceP IA[ABS] [VoiceACT [ERG, EPP] [vP EA [v [VP [V ~~IA~~]]]]]]]]]]]
antecedent

[CP EA[WH] [C[WH, ~~EPP, E~~] [InflP [Infl[ABS] [* VoiceP IA [VoiceACT [ERG, EPP] [vP EA[WH] [v [VP [V IA]]]]]]]]]]]
target clause

I conclude that both the analysis of AF as Exfoliation and the island repair analysis in (123) can account specifically for the data I have been referring to as Active-AF mismatches. However, a unified approach to all the voice mismatches under discussion cannot appeal to island repair by ellipsis, as we will now see.

With that in mind, let us turn to AF-Active mismatches, which I argue involve a configuration like (125):

(124) *AF-Active; IA wh-remnant* (repeated from (79))

A: Xa   xe   ri   ma Juan x-Ø-loq'-o   kotz'i'j.
   EMPH only DET CLF Juan COM-B3S-buy-AF flower
   'Only Juan bought flowers.'
B: Kan qitzij? Ta-b'ij pe chwe   achike kotz'i'j <x-Ø-u-löq'>!
   INT truth   IMP-say DIR PREP.A1S.RN what   flower   COM-B3S-A3S-buy
   'Really? Tell me which flowers!'

(125) *AF-Active mismatch*
a.   *Prior to Exfoliation in the antecedent clause*
[CP [C[FOC, EPP] [InflP [Infl[ABS] [VoiceP IA [VoiceACT [ERG, EPP] [vP EA[FOC] [v [VP [V ~~IA~~]]]]]]]]]]]
antecedent
[CP [C[WH, EPP] [InflP [Infl[ABS] [VoiceP IA[WH] [VoiceACT [ERG, EPP] [vP EA [v [VP [V ~~IA~~]]]]]]]]]]]
target clause

b. *After Exfoliation in the antecedent clause*
[CP [C[FOC, EPP] [InflP [Infl [vP EA[FOC] [v [VP [V IA]]]]]]]]   *antecedent*
[CP [C[WH, EPP] [InflP [Infl[ABS] [VoiceP IA[WH] [Voice ACT[ERG, EPP] [vP EA [v [VP [V ~~IA~~]]]]]]]]]]]
                                                                                    target clause

c. *Focus movement and Sluicing*
[CP EA[FOC] [C[FOC, ~~EPP~~] [InflP [Infl [vP ~~EA[FOC]~~ [v [VP [V IA]]]]]]]]   *antecedent*
[CP IA[WH] [C[WH, ~~EPP, E~~] [InflP [Infl[ABS] [VoiceP IA[WH] [Voice ACT [ERG,~~EPP~~] [vP EA [v [VP [V IA
]]]]]]]]]]]
                                                                                    target clause

On the Exfoliation analysis, sluicing is acceptable in (124)c above because the elided
clause is featurally non-distinct from the antecedent clause, complying with the condition
proposed above (see (89)). We therefore derive that an AF antecedent can mismatch with
an active clause.

Let us now assess whether the island repair approach is feasible for the AF-Active
mismatch. In this case, we would assume that the ellipsis site is in the *AF voice*, matching
in voice with the antecedent. This analysis faces several challenges. First, recall that active
voice must be used when extracting the IA. AF is impossible in such a context:

(126) *AF and object extraction is banned*

    a.   Achike x-Ø-u-tïj   ri a   Juan?
            what   COM-B3S-A3S-eat DET CLF Juan
            'What did Juan eat.'

    b.   *Achike x-Ø-tj-o   ri a   Juan?
            what   COM-B3S-eat-AF DET CLF Juan
            Intended: 'What did Juan eat?'
            Possible reading: 'Who ate Juan?'

An island repair approach necessitates that the examples that I have been calling AF-Active
mismatches actually instantiate AF-AF, but that extraction of the IA using AF *is an island
violation* that is repaired by sluicing. However, it is wholly unclear how the incompatibility
of AF with IA extraction can be taken to be an island violation, regardless of whether the
current proposal of AF as Exfoliation is on the right track. For example, the IA (in the right

configuration) can be the target of successful Agree by Infl^0, with the expected feature-sharing result, in AF clauses (see section 2.4). This suggests that there is no barrier in AF clauses that could create an island violation if the IA is A'-extracted.

An additional observation to make at this point is that sluicing repairs island violations but does not repair all types of deviances. Specifically, preposition stranding violations cannot be repaired by ellipsis, in contrast to island violations (Merchant 2001). The observation that preposition stranding behavior is maintained under ellipsis is called the P-Stranding Generalization (Merchant 2001), and it appears to hold robustly. If a language disallows preposition stranding, it must pied-pipe a preposition along with the *wh*-word in sluicing. Stranding the preposition in the ellipsis site leads to an ungrammatical utterance even under sluicing, in parallel fashion to non-ellipsis cases. Examples from Greek (127)a,b and Basque (127)c,d are given below:[48]

(127)   *P-stranding violation is not repairable by ellipsis*
   a.   I   Anna milise me   kapjon,   all dhe ksero *(me) pjon.
        the Anna spoke with someone but not I.know with who

   b.   *Pjon milise   me?
        who   she.spoke with

   c.   Ana-k   norbait-ekin   hitzegin zuen, baina ez   dakit nor-*(ekin).
        Ana-ERG someone-with talk.to AUX   but   not know who-with

---

[48] Some languages deviate from this expected behavior, including some dialects of Spanish (as already noted by Merchant 2001: 98), Lebanese Arabic (Leung 2014), Indonesian (Fortin 2019), and Polish (Nykiel 2019). However, analyses that do not involve straightforward ellipsis have been proposed to account for the cases where these languages violate the generalization (Rodrigues et.al. 2009; but see Stigliano 2019 for an opposing view). In any case, Kaqchikel is a well-behaved P-stranding language that complies with the P-stranding generalization, so whatever the right analysis is for languages that violate it, this is irrelevant for our purposes here.

d.　　*Nor hitzegin zuen-ekin?
　　　who　talk.to　AUX with　　　　　　　　　　(Merchant 2001: 94-100)

Sluicing, therefore, does not repair *all* violations—i.e., P-stranding violations are incurred

under sluicing as well. An AF-AF plus repair by sluicing approach for the examples I refer

to as AF-Active mismatches must therefore make the case that, whatever goes wrong in

using AF with object extraction, is specifically of the same nature as violations that *are*

repaired by sluicing.[49] I see no straightforward way to argue for such a position, so I

conclude that an island-repair analysis is not a promising alternative for the examples that

I have been calling AF-Active mismatches.

Let us now discuss another alternative analysis for the AF-Active mismatches.[50]

Suppose that there actually exist two movement chains in the target clause. First, the EA

undergoes focus movement to a lower C-level domain, and this extraction triggers AF.

Subsequently, the object *wh*-phrase moves to a position above the extracted EA. Sluicing

then deletes the complement of the higher C^0 head, including the focused EA. The sketch

---

[49] Besides island violations, ellipsis has been argued to repair that-trace effect violations (Perlmutter 1971),

Case Filter violations (Lasnik 2008), and others (see chapter 4 for discussion). Alternative analyses to the

one advanced here would need to assimilate IA extraction from an AF clause with these other types of

violations. I will not assess whether this move is feasible.

[50] In no way does the illustration of this single alternative analysis exhaust the range of possible analyses one

could entertain in order to maintain strict identity between antecedent and sluices in the data discussed. Going

through all possible alternatives would be unfeasible, however, so I merely bring up the challenges that

alternative approaches would need to face in order to account for the issues raised in this chapter: how to

derive voice mismatches in Kaqchikel, while capturing the specific properties of AF (see 2.5).

below provides the schematic structure of the sluice; I use C_2 for the higher C^0 head and C_1 for the lower one. I only show the A'-movements for ease of exposition: [51]

(128) *AF-Active mismatch are actually AF-AF (putative alternative)*

a.   [$_{C2P}$ [$_{C2[WH, EPP, E]}$ [$_{C1P}$ [$_{C1[E, FOC, EPP]}$[$_{InflP}$ [Infl [$_{VoiceP}$ [Voice$_{ACT}$ [$_{vP}$ EA$_{[FOC]}$ [$_v$ [$_{vP}$ [V IA$_{[WH]}$]]]]]]]]]]]]]
target clause

b.   [$_{C2P}$ IA [$_{C2[WH, ~~EPP~~, E]}$ [$_{C1P}$ EA [$_{C1[E, FOC, ~~EPP~~]}$[$_{InflP}$ [Infl [$_{vP}$ ~~EA~~$_{[FOC]}$ [$_v$ [$_{vP}$ [V ~~IA~~$_{[WH]}$]]]]]]]]]]]]
Exfoliation triggered by EA movement

c.   [$_{C2P}$ IA [$_{C2[WH, ~~EPP~~, ~~E~~]}$ [$_{C1P}$ EA [$_{C1[FOC, ~~EPP~~]}$[$_{InflP}$ [Infl [$_{vP}$ EA$_{[FOC]}$ [$_v$ [$_{vP}$ [V IA$_{[WH]}$]]]]]]]]]]]
ellipsis of complement of C_2

Crucially, this analysis assumes that this type of multiple movement is available in general in Kaqchikel. This is in fact correct when extraction targets two arguments: An object *wh*-phrase can be extracted above a focused subject, and this triggers the use of AF. The following is a dialogue in which person A is asking about the different things that a group of people (B, C, D) purchased:[52]

(129) *Wh and focus-extraction of arguments*

A: Achike x-Ø-u-löq'   ri a   Juan?
       what   COM-B3S-A3S-buy DET CLF Juan?
       'What did Juan buy?'

B: Jun *monton*   kotz'i'j.
       one bunch(*Spa*.) flower
       'A bunch of flowers.'

A: Rat, achike x-Ø-a-löq'?
       2SG what   COM-B3S-A2S-buy
       'You, what did (you) buy?'

---

[51] I assume a C domain which is split into at least two projections; see Can Pixabaj & England 2011 for an analysis of the split C domain in closely related K'iche' and Aissen 2017a on the left periphery in Mayan more generally.

[52] Erlewine (2016) reports a variety of multiple extractions that are possible in Kaqchikel, while others are impossible (multiple *wh*-movement, for example). He does not describe this particular configuration, however.

C:	Q'utu'n.
	food
	'Food.'
A:	Achike' loq'oj ja rat x-a-loq'-**o** /*x-Ø-a-löq'? [53]
	what purchases FOC 2SG COM-B2S-buy-**AF** / COM-B3S-A2S-buy
	'What purchases did YOU get?'
D:	Wäy.
	tortillas
	'Tortillas.'

Having established that multiple extraction is possible, we can sketch out what the derivation would be under an approach where multiple movements launch from the ellipsis site. Under this analysis, there is also no voice mismatch—both antecedent and ellipsis site are in the AF voice and therefore lack Voice^0. I will illustrate the logic of the analysis through a derivation where AF is taken to be Exfoliation. Note however that our specific analysis of AF is orthogonal to this alternative analysis, so it is chosen in this context for purely illustrative purposes:

(130) *Alternative analysis: AF-Active mismatches are actually AF-AF*

a. *Prior to Exfoliation in both clauses*

[$_{CP}$ [C$_{[FOC,-EPP]}$ [$_{InflP}$ [Infl$_{ABS}$] [$_{VoiceP}$ IA [Voice$_{ACT \ [ERG, \ \cancel{EPP}]}$ [$_{vP}$ EA$_{[FOC]}$ [v [$_{VP}$ [V $\cancel{IA}$]]]]]]]]]]]   *antecedent*

[$_{C2P}$ [C2$_{[WH, \ EPP, \ E]}$ [$_{C1P}$ [C1$_{[FOC, \ EPP]}$[$_{InflP}$ [Infl$_{[ABS]}$] [$_{VoiceP}$ IA$_{[WH]}$ [Voice$_{ACT \ [ERG, \ \cancel{EPP}]}$ [$_{vP}$ EA$_{[FOC]}$ [v [$_{VP}$ [V $\cancel{IA_{[WH]}}$]]]]]]]]]]]]]]

target clause

b. *After Exfoliation in both clauses*

[$_{CP}$ [C$_{[FOC,-EPP]}$ [$_{InflP}$ [Infl [$_{vP}$ EA$_{[FOC]}$ [v [$_{VP}$ [V IA]]]]]]]]   *antecedent*

[$_{C2P}$ [C2$_{[WH, \ EPP, \ E]}$ [$_{C1P}$ [C1$_{[FOC, \ EPP]}$ [$_{InflP}$ [Infl [$_{vP}$ EA$_{[FOC]}$ [v [$_{VP}$ [V IA$_{[WH]}$]]]]]]]]]]]   *target clause*

c. *Focus movement and Sluicing*

[$_{CP}$ EA$_{[FOC]}$ [C$_{[FOC, \ \cancel{EPP}]}$ [$_{InflP}$ [Infl [$_{vP}$ $\cancel{EA_{[FOC]}}$ [v [$_{VP}$ [V IA]]]]]]]]   *antecedent*

[$_{C2P}$ IA$_{[WH]}$ [C2$_{[WH, \ \cancel{EPP, \ E}]}$ EA$_{[FOC]}$ [C1$_{[FOC, \ EPP]}$ [$_{InflP}$ [Infl [$_{vP}$ EA$_{[FOC]}$ [v [$_{VP}$ [V IA$_{[WH]}$]]]]]]]]]]]

target clause

---

[53] A complex *wh*-phrase appears to be necessary to facilitate the multiple extraction reading. Otherwise, an exclamative interpretation of the *wh*-phrase arises:

(i)	Achike ja rat x-a-loq'-o?!
	what FOC 2SG COM-B2S-buy-AF
	'What?! YOU bought (something)?'

Note that under the alternative analysis sketched out in (130), *strict* syntactic identity is satisfied. If this analysis holds water, then, what I have been calling AF-Active mismatches do not provide evidence for the identity condition predicated on non-distinctness (89).

There is a wrinkle in (130), though. Double extraction is possible only if both extractions target *arguments*. If we attempt to extract an *adjunct* over an EA, double extraction is unacceptable. Consider first an attempt to extract a locative *wh*-phrase above a focused EA:

(131)   *Wh and focus-extraction of locative adjunct and EA*
      a. Ankuchi x-Ø-u-löq'   wi ri   kotz'i'j?
         where   COM-B3S-A3S-buy FP DET flower
         'Where did she buy flowers?'

      b. *Ankuchi ja   ri   xta Maria x-Ø-loq'-o   wi ri   kotz'i'j?
         where   FOC DET CLF Maria COM-B3S-buy-AF FP DET flower
         Intended: 'Where did MARIA buy the flowers?'

      c. *Ankuchi xa   xe   xta Maria x-Ø-loq'-o   wi kotz'i'j?
         where   EMPH only CLF Maria COM-B3S-buy-AF FP flower
         Intended: 'Where did only MARIA buy flowers?'

The same holds with comitative adjuncts:

(132)   *Wh and focus-extraction of comitative adjunct*
      a. Achoj k'in x-Ø-ajo'   wi xta Gilda?
         WH   RN   COM-B3S-dance FP CLF Gilda
         'Who did Gilda dance with?'

      b.*Achoj k'in ja   ri   xta Gilda x-Ø-ajo'   wi?
         WH   RN   FOC DET CLF Gilda COM-B3S-dance FP
         Intended: 'Who did GILDA dance with?'

      c.*Achoj k'in xa   xe   Gilda x-Ø-ajo'   wi?
         WH   RN EMPH only Gilda COM-B3S-dance FP
         Intended: 'Who did only GILDA dance with?'

The fact that these double extractions involving an adjunct are illicit is a challenge to the double movement analysis in (130). Consider once again examples that involve AF in the antecedent and active voice in the target clause. Here, we force the mismatch via the

sprouting of a *wh*-remnant. Recall that adjunct extraction, of which sprouting as seen below

is an instance, is incompatible with AF (see (33)d,f):

(133) *AF-Active mismatch; adjunct wh-remnant*

A: Ja   ri   xta Ana x-Ø-kam-sa-n   ri   aq.
     FOC DET CLF Ana COM-B3S-die-CAUS-AF DET pig
     'ANA killed the pig.'

B: Aw-etama-n   ankuchi x-Ø-Ø-kam-sa-j?
     A2S-know-PERF where   COM-B3S-A3S-die-CAUS-ACT
     'Do you know where she killed it?'

B′: Aw-etama-n   ankuchi <x-Ø-Ø-kam-sa-j>?
     A2S-know-PERF where   COM-B3S-A3S-die-CAUS-ACT
     'Do you know where?'

B″:*Aw-etama-n   ankuchi x-Ø-kam-sa-n?
      A2S-know-PERF where   COM-B3S-die-CAUS-AF
     Intended: 'Do you know where she killed it?'

B‴: *Aw-etama-n   ankuchi ja   ri   xta Ana x-Ø-kam-sa-n?
      A2S-know-PERF where   FOC DET CLF Ana COM-B3S-die-CAUS-AF
     Intended: 'Do you know where ANA killed it?'

A double movement analysis would be faced with the following problem: Whereas double

movement of arguments is licit (129), *wh*-movement of an adjunct over a focused EA is

not (131)-(132). Therefore, a double movement analysis would need to posit that the

unavailability of adjunct *wh*-movement over a focused EA is an *island violation*, and

examples like (133)B′ involve island repair by sluicing. It is unclear to me how we could

argue for such an approach.[54]

---

[54] Furthermore, Howard Lasnik points out to me that some islands cannot be repaired by sluicing with certain

adjunct *wh*-remnants. For instance, Nakao 2009: 59 (citing Lasnik 2005) shows that complex-NP island and

adjunct island violations cannot be repaired by sluicing with a manner *wh*-phrase. The examples below

illustrate this observation:

(i)   John wants to hire [someone who fixes cars **with something**], but I don't know **what**.
                                                                                    complex-NP repair
(ii)   *John wants to hire [someone who fixes cars **in a certain way**], but I don't know **how**.
                                                                                    no complex-NP repair

Nevertheless, let us assume for the sake of argument that *some* repair process is occurring in (133)B′ and there actually is an AF clause in the ellipsis site. Even with this concession, there is another component of a derivation akin to (130) which renders it unfeasible. Notice that in order to ensure that AF is triggered in the sluice, it must be posited that the EA bears focus in the *target clause* (encoded via a [FOC] feature in the derivation), undergoing movement to the left periphery prior to deletion. Therefore, under the analysis in (130), *a focused element needs to be elided*. This is a critical flaw in the putative alternative analysis. Let us see why.[55]

The ellipsis literature has concluded that there are constraints on the information structural status of elements that are elided: Winkler (2016) summarizes the broad consensus that elided elements must be *given* in some sense. We can thus raise the question of whether deleting a focused element is licit in ellipsis at all, since this would be a necessary component of (130). Consider first the dialogue involving several speakers below in English, where VP-ellipsis (VPE) is well-formed:[56]

(134) *Broad focus and VPE*
      A: How fast did John run?

I currently do not have AF-Act mismatches where the *wh*-remnant in the target clause is a manner adverb. Those data would make any analysis taking AF-Active to instantiate AF-AF plus island repair even less feasible. I leave collecting such data for the future.

[55] For other arguments that focused elements cannot be elided, see Weir (2014).

[56] Thanks to Alexander Williams for discussing this type of example with me. An alternative way to characterize the deviance of (135) is that the question-answer congruence between C and D is wrong—using VP-ellipsis in D signals broad focus on the answer, which is inadequate given C. I interpret this as being notionally and extensionally equivalent to my interpretation of the deviance—narrowly focused elements cannot be elided.

B: He ran very fast.
C: What about Mary?　　　　(*broad focus question*)
D: She did <run very fast> too.　(*VPE well-formed*)
D′: She ran very fast too.

In the dialogue above, speaker B's remark renders VP ellipsis well-formed across C's question, which is a *broad* focus question answered by speaker D. However, if we modify speaker C's question such that it forces *narrow* focus on the low manner adjunct, VP-ellipsis becomes infelicitous:

(135)　*Narrow focus of V-internal elements and VP-ellipsis*
A: How fast did John run?
B: He ran very fast.
C: How fast did Mary run?　　(*narrow focus question*)
D: #She did <run very fast> too. (*VPE ill-formed*)
D′: She ran very fast too.

In a nutshell, attempting to elide an element that is focused is not possible, so any analysis that requires the ellipsis of a focus-marked element to get off the ground is on the wrong track.[57] I therefore reject (130) as a possible alternative analysis of AF-Active mismatches.

---

[57] The ban on eliding focused elements is actually more subtle. A cleft, for instance, can occur inside an ellipsis site:

(i) Mary said that it was John who danced salsa, and Robin did <say that **it was John** who danced salsa> too.

Note that in (i) however, the *it*-cleft is inside an embedded clause. This seems relevant in determining where focused elements can occur inside the ellipsis site (i.e., in the embedded CP in the ellipsis site, as in (i)). I leave further exploration of these patterns for future work, noting that the alternative analyses proposing AF-AF, discussed in the main text, are analogous to (135) (clausemate focus), rather than (i) (focused element in a separate embedded clause).

Let us return to the last mismatch we described in subsection 2.2.3.4, the Passive-AF mismatch. Observe the relevant type of data once again.[58]

(136) *Passive-AF mismatch*

    a. Ru-ma jun winäq, x-e-k'ay-ïx   aq pa   k'ayib'äl. Ri a   Pedro
        A3S-RN a   person COM-B3PL-sell-PAS pig PREP market   DET CLF Pedro
        x-Ø-u-kut-uj   achike winäq <x-e-k'ay-in   aq >?
        COM-B3S-A3S-ask-ACT which man   COM-B3P-sell-AF pig
        'Pigs were sold by some person in the market. Pedro asked which person.'

    b. Wawe' x-e-k'ay-ïx   äk'   ru-ma jun ixöq.   Aw-etama-n
        here   COM-B3PL-PAS chicken A3S-RN a   woman. A2S-know-PERF
        achike ixöq <x-e-k'ay-in   äk' > ?
        which woman COM-B3P-sell-AF chicken
        'Chickens were sold here by some woman. Do you know which woman?'

    c. Ri   aq x-Ø-kam-is-äx   ru-ma jun winäq.   Aw-etama-n
        DET pig COM-B3S-die-CAUS-PAS A3S-RN a   person   A2S-know-PERF
        achike winäq <x-Ø-kam-sa-n   ri   aq t_1> ?
        which person COM-B3S-die-CAUS-AF DET pig
        'The pig was killed by some person. Do you know which person?'

It should be clear how the examples in (136) follow directly from the syntactic identity condition in (89) and our analysis of AF as Voice Exfoliation: an antecedent specified for VoicePASS can mismatch with an AF sluice, because AF instantiates a configuration without a Voice layer. In other words, the antecedent and sluice are featurally non-distinct in examples (136)a-c. Ellipsis is therefore allowed.

---

[58] For this example specifically, one of my consultants accepted (136)c, but showed a preference for an alternative version of the sentence where a determiner follows the *wh*-remnant (the spading-like construction discussed in section 2.2.2.5):

    (i) Ri   aq x-Ø-kam-is-äx   ru-ma jun winäq. Aw-etama-n   achike winäq ri'?
        DET pig COM-B3S-die-CAUS-PAS A3S-RN a   person A2S-know-PERF which person DET
        'The pig was killed by some person. Do you know which person it was?'

The properties of spading will be explored more thoroughly in future work.

As we have done before, let us now assess the feasibility of an alternative analysis in which there is no voice mismatch. Under such an approach, the examples in (136)a-c necessarily involve a voice match, so the sluices would need to contain *passive voice*. This would mean that the subject *wh*-remnant in the target clause was extracted from a passive. Furthermore, it would imply that the *wh*-phrase was extracted and *stranded the relational noun* which necessarily introduces the thematic agent of a passive (specifically, the complements of *ruma* in (136)a-c). I use the labels A (agent) and T (theme) below:

(137) *Passive-AF mismatches are actually Passive-Passive (putative alternative)*
a.   V-PASSIVE T [RN A$_{INDEFINITE}$]   *antecedent*
b.   A$_{WH1}$ <V-PASSIVE T [RN t$_1$]>   *target clause*

An immediate problem with the approach in (137) is that it necessitates that relational noun stranding be possible in Kaqchikel. In other words, Kaqchikel would need to allow P-stranding for the analysis to be viable. This is not the case, however, as shown by the comitative adjunct *wh*-question in (138), the benefactive adjunct *wh*-question in (139), and the reason *wh*-question in (140) below:

(138) *Relational noun stranding is impossible in Kaqchikel; comitative adjunct wh-question*

a. Rat k'o   achoj k'in x-a-b'e   pa   Armita.   Ta-b'ij pe chwe   **[achoj**
  2S EXIST WH   RN COM-B2S-go PREP Guatemala.City IMP-say DIR PREP.RN.A1S **WH**
  k'in]$_1$ x-a-b'e   pa   Armita   **t$_1$** !
  RN   COM- B2S-go PREP Guatemala.City
  'You went with someone to Guatemala city. Tell me who you went to G. C. with!'

b. Rat k'o achoj k'in xab'e pa Armita. Tab'ij pe chwe **achoj k'in$_1$** <xab'e pa Armita **t$_1$**>!

c. *Rat k'o achoj k'in xab'e pa Armita. Tab'ij pe chwe **achoj**$_1$ <xab'e pa Armita **k'in t$_1$**>!

(139) *Relational noun stranding is impossible in Kaqchikel; benefactive adjunct wh-question*

a. Rat k'o   achoj chin x-Ø-a-löq'   re kotz'i'j re'. Ta-b'ij pe chwe   **[achoj**
   2S EXIST WH   RN COM-B3S-A2S-buy DEM flower DEM IMP-say DIR PREP.RN.A1S **WH**
   chin]₁ x-Ø-a-löq'   re kotz'i'j re' **t₁**!
   RN   COM-B3S-A2S-buy DEM flower DEM
    'You bought these flowers for someone. Tell me who you bought these flowers for!'

b. Rat k'o achoj chin xalöq' re kotz'i'j re'. Tab'ij pe chwe **achoj chin₁** <xalöq' re kotz'i'j re' **t₁**>!

c. *Rat k'o achoj chin xalöq' re kotz'i'j re'. Tab'ij pe chwe **achoj** <xalöq' re kotz'i'j re' **chin t₁**>!

(140) *Relational noun stranding is impossible in Kaqchikel; reason adjunct wh-question*

a. Rat k'o   achike ru-ma   x-Ø-a-löq'   re kotz'i'j re'. Tab'ij pe chwe
   2S EXIST WH   A3S-RN COM-B3S-A2S-buy DEM flower DEM IMP-say DIR PREP.RN.A1S
   [achike ru-ma]₁ x-Ø-a-löq'   re kotz'i'j re' **t₁**!
   WH   **A3S-RN** COM-B3S-A2S-buy DEM flower DEM
    'You bought these flowers for some reason. Tell me why you bought these flowers!'

b. Rat k'o achike ruma xalöq' re kotz'i'j re'. Tab'ij pe chwe **achike ruma₁** <xalöq' re kotz'i'j re' **t₁**>!

c. *Rat k'o achike ruma xalöq' re kotz'i'j re'. Tab'ij pe chwe **achike₁** <xalöq' re kotz'i'j re' **ruma t₁**>!

Put simply, Kaqchikel complies with Merchant 2001's P-stranding generalization—the language disallows stranding of relational nouns in general and inside ellipsis sites: given (138)-(140), the configuration in (137) would involve eliding a stranded relational noun—an otherwise impossible structure in Kaqchikel. As we discussed earlier, a preposition stranding violation cannot be repaired under ellipsis, in contrast to island violations (Merchant 2001, Stjepanovic 2012, Larson & Hornstein 2013). I know of no language where P-stranding is extraordinarily allowed in ellipsis only in order to avoid a violation of the identity condition (in this case, a Passive-AF mismatch). Those languages that are *prima facie* exceptions to the P-stranding generalization are exceptions to it across the board; i.e., regardless of whether a violation of voice identity is at issue in the relevant

examples. Therefore, the derivation in (137) is illicit—such an analysis would require Kaqchikel to allow relational noun stranding violations (i.e., P-stranding violations) to be repaired by ellipsis, contrary to fact.

There is an additional, serious problem for the analysis in (137) that we are currently evaluating. The *wh*-remnant in the examples in (136), *achike*, is not the *wh*-phrase that is used for asking questions about the demoted agent of a passive. If *achike* is used alongside the *uma* RN, only a reason adjunct reading is possible:

(141) *Questioning the demoted agent of a passive*
      Achike ru-ma x-Ø-kam-is-äx   ri   aq?
      WH   A3S-RN COM-B3S-die-CAUS-PAS DET pig
      #'By who was the pig killed?' / 'Why was the pig killed?'   (*all speakers*)

For speakers who are able to construct a question targeting the demoted agent of a passive, the question word that is used in such a context is *achoq*, not *achike*—which is the question word we see in examples (136)a-c and all Passive-AF mismatches more generally:[59]

(142) *Questioning the demoted agent of a passive*
      Achoq ru-ma x-Ø-kam-is-äx   ri   aq?
      WH   A3S-RN COM-B3S-die-CAUS-PAS DET pig
      'By who was the pig killed?'   (Lolmay Pedro García Matzar p.c.)

The evidence points to the following conclusion, then: There is no straightforward way that the derivation of the examples in (136) involves a passive Voice[0] in the sluice. Such an

---

[59] A disclaimer is necessary here. The data in (142) was provided to me by Lolmay Pedro García Matzar, Kaqchikel linguist and speaker of the San Andrés Semetabaj dialect, who authored the reference grammar mentioned throughout the chapter. I have been unable to replicate the judgement for the example in (142) with any of my three consultants in Patzún—none of them accept the question as involving the demoted agent of a passive and they all interpret the question as a benefactive '*Who was the pig killed for*?' See section 2.5 for some discussion.

analysis would face two challenges: (i) Explaining why a P-stranding violation is repaired under ellipsis, extraordinarily and in a way that is not attested elsewhere in the grammar of Kaqchikel, and (ii) why an unexpected *wh*-phrase surfaces as a remnant under sluicing. In contrast, the proposed analysis of AF as Voice Exfoliation, paired with the syntactic identity condition in (89), can derive AF-Passive mismatches, in addition to the other Voice mismatches in Kaqchikel.

To summarize then, we have observed that (i) Active-AF, (ii) AF-Active, and (iii) Passive-AF voice mismatches in Kaqchikel sluicing can be handled in a unified manner by the proposed syntactic identity condition. Putative alternative analyses were assessed, and we observed that they face significant challenges. Therefore, I conclude that the approach advanced here is on the right track. The analysis, however, has additional advantages beyond capturing the set of well-formed and ill-formed voice mismatches. In the next section, I discuss how the Exfoliation approach to AF can derive many of the idiosyncratic properties of AF.

2.4 AF as Exfoliation of Voice^0: Additional consequences

In this section, I show how the proposed analysis of AF as Exfoliation of Voice^0 has the advantage of explaining many of the properties of AF that have puzzled researchers in the past. The properties of AF we will discuss are the following:

(143)   *Properties of AF*
      a. ERG agreement does not surface.
      b. Only ABS agreement surfaces.
      c. ABS agreement is controlled either by the EA or the IA, depending on an accessibility hierarchy 1,2 >3PL >3S
      d. AF cannot be used if both arguments are local persons (i.e., 1st and 2nd person).

In a sense, this section serves the following purpose: I intend to show that while the analysis advanced here accounts for the (otherwise) puzzling Voice mismatches in Kaqchikel sluicing, it also has desirable consequences beyond that empirical domain. In particular, the idiosyncratic properties of AF have proven a challenge for many theorists over the years (see Aissen 2017b for a summary), but the proposed analysis has a way of handling these properties as a consequence of the removal of Voice via Exfoliation. Therefore, the analysis has virtues within the narrower domain of analyzing a construction that is particular to Kaqchikel (and other Mayan languages). At the end of this section, I will also discuss why AF cannot be used when the internal argument is a reflexive or an extended reflexive—in brief, I will argue that in such a configuration, the structural description of Exfoliation is not met, because the internal argument does not move to the edge of VoiceP.

The one thing I will not explore here is how analyses of AF proposed elsewhere could handle the sluicing data (for example, Stiebels 2006, Aissen 2011, Coon et al. 2014, Assman et al. 2015, Erlewine 2016, Baier 2018, Coon et al. 2021). I leave such a comparison for the future. However, I do note the following: none of the works where these proposals are put forth discuss the ellipsis phenomenon that has been our focus in this chapter.

2.4.1 No ERG agreement

While neither argument is demoted in AF, in the sense of being omitted or being introduced as the complement of a RN, Set A agreement morphology does not appear. Only a single agreement morpheme surfaces, from the Set B paradigm:

(144)   *No ERG agreement in AF*

     a.   Ri a   Juan x-Ø-u-tïj   nu-way.
          DET CLF Juan COM-B3S-A3S-eat A1S-tortilla
          'Juan ate my tortilla.'

     b.   *Achike x-Ø-u-tj-o   nu-way?
          who   COM-B3S-A3S-eat-AF A1S-tortilla
          Intended: 'Who ate my tortilla?'

     c.   Achike x-Ø-tj-o   nu-way?
          who   COM-B3S-eat-AF A1S-tortilla
          'Who ate my tortilla?'

Under the proposed analysis of AF, this is expected. Set A morphology arises in an active clause via clitic-doubling resulting from an Agree relation between Voice and the EA. In AF, the Voice projection has been deleted by Exfoliation. As a result, set A morphology is deleted:

(145)   *Extraction restriction on the EA*

    [C$_{[WH]}$ [Infl [$_{VoiceP}$ IA [Voice$_{ACT\,[ERG]}$ [$_{vP}$ EA$_{[WH]}$ [v [$_{VP}$ [V ~~IA~~]]]]]]]]

                              Syntactic ergativity

(146)   *AF is Exfoliation*

a. [C$_{[WH]}$ [Infl [$_{VoiceP}$ IA [Voice$_{ACT\,[ERG]}$ [$_{vP}$ EA$_{[WH]}$ [v [$_{VP}$ [V ~~IA~~]]]]]]]]

                              Target of Exfoliation

b. [C$_{[WH]}$ [Infl [$_{vP}$ EA$_{[WH]}$ [v [$_{VP}$ [V IA]]]]]]   *Exfoliation applies to Voice, ERG deleted*

In other words, a consequence of the deletion of the Voice projection is that the Set A (ERG) agreement clitic is deleted as well. We therefore derive the disappearance of Set A agreement in AF.

2.4.2 Only ABS agreement surfaces; Infl^0 can Agree with either IA or EA

The analysis advanced here can handle two additional properties of AF: (i) Only set B morphology surfaces and (ii) set B agreement is controlled by either the EA or IA.

Let us first lay out the agreement pattern in AF.[60] In all of the AF examples we have observed so far, both IA and EA have been 3S—thus, it has not been surprising that we have encountered only the 3S Set B marker, which is null. However, set B agreement in AF is governed by the descriptive hierarchy given below. If either argument is a local person, then that argument will control agreement; if either (or both arguments) are 3P, then 3P agreement will surface; if both arguments are 3S, then the aforementioned null marker occurs (García Matzar & Rodríguez Guaján 1997, Preminger 2014).

(147) *Set B agreement controllers in AF* (Preminger 2014: 21)
       1st/2nd person > 3rd person plural > 3rd person singular

A sample of relevant data showing the effect of the hierarchy is given below. The argument pairs in the AF sentence below are 1S and 3S. Note that 1S set B agreement surfaces in both examples, regardless of whether the local person is an EA (148)a or an IA (148)b:

(148) *Agreement hierarchy effect in AF: 1S > 3S*
       a. Ja   yïn x-in-/*Ø-ax-an   ri   achin.
          FOC 1S COM-A1S-/*B3S-hear-AF DET man
          'It was me that heard the man.'
       b. Ja   ri   achin x-in-/*Ø-ax-an   yïn.
          FOC DET man   COM-A1S-/*B3S-hear-AF 1S
          'It was the man that heard me.'   (adapted from Preminger 2014: 57)

By hypothesis, in both examples above, Exfoliation has deleted the VoiceP projection. Given the Exfoliation of Voice, the copy of IA residing in SpecVoiceP is also deleted. As a result, the tail of the clitic doubling chain <Infl^0, IA> is deleted:

(149) *AF*
a. [C$_{[WH]}$ [Infl [$_{VoiceP}$ IA [Voice$_{ACT [ERG]}$] [$_{vP}$ EA$_{[WH]}$ [v [$_{VP}$ [V ~~IA~~]]]]]]]]
                                                                    Target of Exfoliation

---

[60] For an in-depth discussion of agreement in K'ichean AF, see Preminger (2014). I follow the spirit of Preminger's analysis here.

b. [C$_{[WH]}$ [Infl [$_{vP}$ EA$_{[WH]}$ [v [$_{VP}$ [V IA]]]]]]

Exfoliation applies to Voice, higher copy of IA deleted

Let us stipulate that the severing of the clitic-doubling chain requires Infl^0 to probe again. Given that there is no longer any phase boundary between Infl^0 and either argument, Infl^0 can target whichever argument bears the relevant feature and establish a feature sharing relation.[61] Assume that Infl^0 probes for [PART(ICIPANT)]. In (148)a above, the EA argument is 1S, bearing a [PART] feature targeted by Infl^0. Infl^0 probes its domain and finds the EA argument, resulting in a successful Agree relation. As a result, a 1S set B morpheme surfaces on the verbal stem. In (148)b, the IA is 1S, bearing [PART]. Infl^0 probes its domain and finds the IA, resulting in a successful Agree relation and the 1S set B marker that we observe. Note that in (148)b, the 3S EA does not intervene in the Agree relation because a 3S bears no feature whatsoever (Preminger 2014). For completeness, note than when there is only one argument for the Infl^0 probe to inspect (e.g., the internal argument in active transitives and unaccusatives), the probe is relativized to the same feature as in the AF

---

[61] Having Infl^0 probe following Exfoliation might raise concerns of cyclicity, since Exfoliation was triggered by the [WH] needs of C^0, which were impossible to satisfy otherwise. If all operations at the Infl level must precede all operations at the C level, post-Exfoliation probing by Infl seems to be problematic. However, if we assume that all operations within a phase occur simultaneously (Chomsky 2008), then this apparent countercyclicity disappears; see section 2.5.3 for how this can handle an Exfoliation derivation where Spec,VoiceP survives. Note, furthermore, that this sort of phase-bounded counter-cyclicity has been independently shown to be necessary for φ-agreement (see Holmberg & Hróarsdóttir 2003).

configuration. However, since there is only one potential goal in these configurations, the omnivorous nature of the probe is not detectable, unlike in AF clauses.[62]

To recapitulate, notice that the EA can only enter into an Agree relation with $Infl^0$ in (149)a because Exfoliation removed the Voice phase, severing the clitic-doubling chain <Voice, EA>. Had Exfoliation not applied, $Infl^0$ would not have probed past $Voice^0$ and entered into the relation with the EA. Even if it were able to do so, it would not be able to target the EA because the latter would still be the lower copy in its own clitic-doubling chain. It is for this reason that the EA can only be targeted by $Infl^0$ (giving rise to set B agreement) in AF, which involves the removal of $Voice^0$—only in such a configuration is the EA a non-clitic-doubled nominal that can be targeted by $Infl^0$.

2.4.3 Incompatibility of AF with two local arguments

When both arguments are local persons, AF is banned:

(150) *Two local persons in AF is banned*
      a. *Ja   rat x-in-/at-/Ø-ax-an   yïn.
          FOC 2S COM-B1S-/B2S-/B3S-hear-AF 1S
          Intended: 'It was you that heard me.'
      b. *Ja   yïn x-in-/at-/Ø-ax-an   rat.
          FOC 1S COM-B1S-/B2S-/B3S-hear-AF 2S
          Intended: 'It was me that heard you.'   (adapted from Preminger 2014: 66)

---

[62] The precise characterization of $Infl^0$ probing in Kaqchikel is more complex. As Preminger (2014) argues, it is necessary to deconstruct the probe for ABS agreement in AF into two separate probes—first, a person head π^0 probes, followed by a number $\#^0$ probe (for discussion on whether this involves separate heads, as opposed to two probes on the same head, I refer the reader to Preminger's work). This decoupling is necessary in order to account for the agreement configuration in AF where 3P wins over 3S. Since our purpose here is not to derive the entire space of agreement possibilities in AF, I do not delve into this in more detail.

I assume following Preminger (2014) that this type of configuration is banned because arguments bearing a [PART] feature need to be licensed in the course of the derivation.

(151) *Person Licensing Condition* (Béjar & Rezac 2003 *apud* Preminger 2014)
       Interpretable $1^{st}/2^{nd}$ features must be licensed by entering into an Agree relation with an appropriate functional category.

Under my account, we must understand (151) as a representational, not derivational requirement. The effect of Exfoliation of $Voice^0$ is as follows: In a regular active clause, an EA bearing [PART] enters into an Agree relation with $Voice^0$, a relation that licenses the argument. However, following Voice Exfoliation, the relation between $Voice^0$ and such an EA is broken, since the head of the clitic-doubling movement chain <$Voice^0$, EA> is deleted. The EA thus becomes available once more for probing by another head. If the EA is the only argument bearing [PART], then entering into an Agree relation with $Infl^0$ will satisfy (151).

Notice though that a problem arises if Exfoliation applies and *both* the EA and the IA bear a [PART] feature. Both arguments require licensing, but only one can be targeted by $Infl^0$. In such a configuration, the derivation crashes, since one of the arguments is left unlicensed:

(152) *AF with two arguments bearing [PART]*

a. [$C_{[A']}$ [Infl [$_{VoiceP}$ $IA_{[PART]}$ [$Voice_{ACT[ERG]}$ [$_{vP}$ $EA_{[A', PART]}$ [v [$_{VP}$ [V ~~$IA_{[PART]}$~~]]]]]]]]   *Exfoliation*

b. [$C_{[A']}$ [Infl [$_{vP}$ $EA_{[A', PART]}$ [v [$_{VP}$ [V **$*IA_{[PART]}$**]]]]]]   *$IA_{[PART]}$ unlicensed*

To summarize: the proposed analysis can make sense of why a configuration with two local person arguments bans AF. If both arguments bear [PART] and Exfoliation applies, a single

128

probe Infl^0 can only license *one* argument. As per the PLC, such a derivation leads to a crash.[63]

2.4.4 Summary of properties of AF that derive from Exfoliation

I started this section by noting that AF exhibits several properties that are a challenge for any analysis of the syntax of the construction:

(153)   *Properties of AF clauses*
   a. ERG agreement does not surface.
   b. Only ABS agreement surfaces.
   c. ABS agreement is controlled either by the EA or the IA, depending on an accessibility hierarchy 1,2 >3PL >3S
   d. AF cannot be used if both arguments are local persons (i.e., 1st and 2nd person).

I have shown that the proposed analysis of AF as Exfoliation accounts for these properties. In the next subsection, I discuss one final property, before turning to some remaining issues in 2.5.

2.4.5 Incompatibility of AF with a reflexive/ extended reflexive IA

AF cannot be used when the IA is a reflexive, an observation that goes at least back to Mondloch 1981 for K'iche', another K'ichean language exhibiting syntactic ergativity (see Aissen 2017b for discussion of syntactic ergativity across the family; see also Ordóñez 1995, Aissen 2011, Coon et al. 2014, Coon et al. 2021, Burukina 2019 for the lifting of the extraction constraint when the IA is a reflexive in different Mayan languages).

---

[63]Note that Preminger 2014 and Coon et al. 2021's analyses also derive the AF-person restriction. A question arises regarding which kind of construction *is* used to focus the external argument in Kaqchikel when both arguments are local persons. In this configuration, *active voice* is licit (see Erlewine 2016), as is the AP_OBL. I set aside this issue here, noting that it requires explicit discussion in future work.

Additionally, Henderson & Coon (2018) note that an extended reflexive IA is incompatible with AF in Kaqchikel as well. An extended reflexive is a possessed nominal whose possessor is bound by the EA (see Aissen 1999):[64]

(154)   *AF is incompatible with a reflexive IA*
    a. **Achike** x-Ø-u-chäy   **r-i'**?
       who   COM-B3S-A3S-hit **A3S-REFL**
       'Who hit himself?'

    b. ***Achike** x-Ø-ch'ey-o   **r-i'**?
       who   COM-B3S-hit-AF **A3S-REFL**
       Intended: 'Who hit himself?'   (Burukina to appear)

(155)   *AF is incompatible with an extended reflexive IA*
    Achike$_1$ x-Ø-u-k'äm   pe **r$_1$**-ixayil pa   nimaq'ij?
    what   COM-B3S-A3S-bring DIR **A3S**-wife PREP party
    'Who brought his (own) wife to the party?'   (Henderson & Coon 2018)

Two options to account for the well-formed status of (154)a and (155) arise here. The first is to follow Burukina 2019 in proposing that reflexives enter into a relation with a special Refl^0 voice head (see also Labelle 2008) and assume, further, that this head is not a phase.[65] As a result, the structural description of Exfoliation is never met—the external argument is able to be targeted by the probe on C and move. A second alternative is to tie the movement of the internal argument to Case licensing, such that said argument only moves to the edge of VoiceP in order to be targeted and licensed by Infl^0. If reflexives and extended reflexives are licensed in an extraordinary way (i.e., via incorporation or pseudo-

---

[64] Henderson & Coon 2018 do not give an example showing that AF is *impossible* with an extended reflexive—they only show, via (155), that using active voice is possible here.

[65] Given that Burukina assumes a slightly different clause structure than I do—i.e., she assumes [*v*P [VoiceP [VP]]]—I leave for the future a thorough evaluation of how the proposed analysis here and hers can be rendered fully compatible.

incorporation; see Coon et al. 2014), then movement does not proceed and the structural description of Exfoliation is not met—thus, the external argument is free to extract in this configuration, since the internal argument never occupied the edge of the phase.

2.5 Remaining issues in Kaqchikel

This section is devoted to describing some remaining issues, shown below. I leave a full discussion and solutions for the puzzles raised here for future work:

(156) *Remaining issues*
    a. The AP$_{OBL}$ voice
    b. All possible voice combinations under sluicing
    c. The fate of specifiers after Exfoliation

2.5.1 The AP$_{OBL}$ voice

Recall from section 2.2 that there is an antipassive Voice which I deemed AP$_{OBL}$. To remind the reader, the thematic object of a transitive is introduced in this voice in an oblique phrase headed by a relational noun *ichin*. The voice suffix arising in this construction is identical to the suffix that arises in AF, for derived and underived stems alike. An example is given below once more.[66]

(157) *Grammatical AP$_{OBL}$ with a focused EA*
    Ri ixöq   x-Ø-kem-o   r-ichin ri po't.
    DET woman COM-B3S-weave-APO A3S-RN DET güipil
    'THE WOMAN wove the güipil.'   (adapted from Heaton 2017: 345)

A question that arises naturally in the context of our investigation is whether the AP$_{OBL}$ can mismatch with other voices under sluicing. As shown below, it can:

---

[66] While Heaton's translation in (157) does not indicate it, it is clear from her discussion in the text that this example involves subject focus.

(158) *APobl-Active (non-sluice)*

A: Xta Andrea x-Ø-Ø-kam-sa-j   jun chiköp!
   CLF Andrea COM-B3S-A3S-die-CAUS-ACT one animal
   'Andrea killed an animal!'

B: Na! Ja   ri   xta Maria **x-Ø-Ø-kam-sa-n**   r-ichin ri   chiköp.
   no   FOC DET CLF Maria **COM-B3S-die-CAUS-APO** A3S-RN   DET animal.
   'No! MARIA killed the animal.'

A: A… Aw-etama-n   achike chiköp **x-Ø-Ø-kam-sa-j**?
   oh   A2S-know-PRF what   animal **COM-B3S-A3S-die-CAUS-ACT**
   'Oh… Do you know what animal she killed?

B: Ja,   w-etama-n.
   yes A1S-know-PRF.
   'Yes, I know.'

(159) *APobl-Active mismatch*

A: Xta Andrea x-Ø-Ø-kam-sa-j   jun chiköp!
   CLF Andrea COM-B3S-A3S-die-CAUS-ACT one animal
   'Andrea killed an animal!'

B: Na! Ja   ri   xta Maria **x-Ø-kam-sa-n**   r-ichin ri   chiköp.
   no   FOC DET CLF Maria **COM-B3S-die-CAUS-APO** A3S-RN DET animal.
   'No! MARIA killed the animal.'

A: A… Aw-etama-n   achike chiköp <**x-Ø-Ø-kam-sa-j**>?
   oh   A2S-know-PRF what   animal   **COM-B3S-A3S-die-CAUS-ACT**
   'Oh… Do you know what animal she killed?

B: Ja,   w-etama-n.
   yes A1S-know-PRF.
   'Yes, I know.'

We see in the dialogue in (159) above that sluicing is licit even though there is a mismatch between AP$_{OBL}$ in the antecedent (A's first remark) and active voice in the sluice (A's second remark). We therefore seem to have a case where an antipassive can mismatch with active voice. Recall also, that as shown in section 2.2.3.1, there is a different antipassive in Kaqchikel (the AP) that *cannot* mismatch with active voice. These data could raise an issue for our analysis, then: if a bona fide Voice^0 underlies the oblique antipassive in the syntax,

then we would predict that an AP$_{OBL}$-Active mismatch like in (159) would be unacceptable, since it violates the identity condition that I have been defending thus far:

(160) *Syntactic identity in ellipsis*
       The antecedent and material properly contained within the ellipsis site must be featurally non-distinct.

However, the worry that might be raised by these data is only an illusion arising from using the term "antipassive" for the construction in (157). What I will argue here is that the syntax of the AP$_{OBL}$ is identical to AF. In other words, the mismatch in (159) complies with featural non-distinctness, since the AP$_{OBL}$ instantiates the removal of Voice via Exfoliation, just like AF does. To my knowledge, an analysis of the syntax of the Kaqchikel AP$_{OBL}$ is still awaiting.[67] Here, I will attempt to unify the analysis of AF with the AP$_{OBL,}$ primarily to account for the sluicing data under consideration. We will see, though, that there are independent reasons to unify the syntax of AF and the AP$_{OBL}$.

Let us start with the following observation, made by Heaton (2017), that there is something amiss with calling the AP$_{OBL}$ a run-of-the-mill antipassive:

(161) *Heaton's observation*
       The syntactic distribution of AF is identical to the syntactic distribution of the AP$_{OBL}$.   (Heaton 2017: 342-52)

In other words, the AP$_{OBL}$ is every bit as Last Resort as AF. The AP$_{OBL}$ can only be used if the EA is A'-extracted. In a sense, then, calling this voice an "antipassive" is a misnomer,

---

[67] For example, Aissen (2011) analyzes AF in closely related K'ichee', but sets aside the AP$_{OBL}$. As Nora England reminds me, my proposal here attempts to formalize and unify the syntax of AF and AP$_{OBL}$, which has been at the forefront of many Mayanists' mind when they take the two constructions as being the same in a deep sense (see e.g. Aissen 2017b).

since this Last Resort flavor is not a general property of antipassives, which can be used whenever the object has the right properties (see Polinsky 2017):[68]

(162) *Grammatical AP_OBL with a focused EA (repeated from (157))*

Ri ixöq x-Ø-kem-o r-ichin ri po't.
DET woman COM-B3S-weave-APO A3S-RN DET güipil
'THE WOMAN wove the *güipil*.'

(163) *Ungrammatical AP_OBL with a non-focused EA*

*X-Ø-kem-o ri ixöq r-ichin ri po't
 COM-B3S-weave-APO DET woman A3S-RN DET güipil
Intended: 'The woman wove the *güipil*.' (adapted from Heaton 2017: 345)

The AP_OBL is incompatible with A'-extraction of any element other than the EA. In this way, the AP_OBL is parallel once more to AF. Contrast the examples given in (164)-(165) which illustrate this fact. Furthermore, observe that, in contrast to AP_OBL, adjunct extraction is compatible with AP. The data show then that the AP_OBL voice is unlike antipassives, and akin to AF:

(164) *AP_OBL with EA A'-extraction licit*

a. Achike x-Ø-loq'-o r-ichin ri a-ch'ich''?
   who COM-B3S-buy-APO A3S-RN DET A2S-car
   'Who bought your car?'

b. Ja ri xta Ana x-Ø-tz'et-o r-ichin xta Maria.
   FOC DET CLF Ana COM-B3S-see-APO A3S-RN CLF Maria
   'ANA saw Maria.'

(165) *AP_OBL with adjunct wh-extraction illicit*

*Ankuchi x-Ø-loq'-on (wi) r-ichin ri kotz'i'j?
 where COM-B3S-buy-APO FP A3S-RN DET flower
 Intended: 'Who bought the flowers?'

---

[68] Recall, in contrast, that the AP (i.e., the absolute antipassive) can be used in contexts where there is no A'-extraction of the EA; see (74)a.

(166) *AP with adjunct locative extraction*

Akuchi' x-at-loq'-on   wi?
where   COM-B2S-buy-AP FP
'Where did you go buying (something)?'   (Mendes & Ranero 2021)

(167) *AP with instrumental adjunct extraction*

R-ik'in jun xik x-i-tzib'-an   wi.
A3S-RN a   pen COM-B1S-write-AP FP
'With a pen I wrote.'   (Mendes & Ranero 2021 citing Silberman 1995)

In a nutshell, AF and AP$_{OBL}$ have the same distribution; AP$_{OBL}$ and AP have a different

distribution.[69] While the passive (as well as the AP) can be used in different syntactic

frames, AP$_{OBL}$ is limited to A'-extraction of the EA, just like AF.

I propose therefore that AP$_{OBL}$ has the following derivation. The IA is merged as an

oblique phrase from the start. The derivation proceeds as we observed with AF: The IA

---

[69] Erlewine 2016, citing McKenna Brown et.al. 2006, provides the following example of an apparent oblique

antipassive without A'-extraction. Notice that the theme is introduced below by the expected relational noun

ichin:

(i)   Y-i-tz'et-on   (aw-ichin).
       INC-B1S-see-AP A2S-RN
       'I see you.'   (adapted from Erlewine 2016)

The status of this data point is unclear to me, in the context of our investigation. Heaton 2017 (p. 379-383)

notes that Patzún speakers are conservative in comparison to other Kaqchikel speakers, since they have a

clear form contrast between the form of the AP (-Vn) from the AP$_{OBL}$ (-o/-u) with CVC roots, something

that I have observed as well in my work (see section 2). Notice, then, that the form of the antipassive suffix

in (i) is –Vn, which is the form for the AP in Patzún Kaqchikel, not the AP$_{OBL}$. This example must therefore

not be representative of a Patzún Kaqchikel speaker. Additionally, McKenna Brown et.al 2006 is a

descriptive grammar that is intended primarily for pedagogical purposes. In other words, the Kaqchikel found

therein might be prescriptive or anachronistic. Therefore, the relevance and status of examples like (i) is

unclear, since we are analyzing the synchronic grammar of Patzún Kaqchikel speakers.

moves to SpecVoiceP and the EA cannot extract. As a result, Exfoliation applies. Below, I label the oblique IA in the AP$_{OBL}$ as a PP for ease of exposition, even though in Mayan-specific terms, we are dealing with a relational noun (RN) structure:

(168) *Derivation of the AP$_{OBL}$*

    a. [$_{CP}$ [C$_{[WH, EPP]}$ [$_{InflP}$ [Infl$_{[ABS]}$ [$_{VoiceP}$ PP [Voice$_{active\ [ERG, \sout{EPP}]}$ [$_{vP}$ EA$_{[WH]}$ [v [$_{VP}$ [V ~~PP~~]]]]]]]]]]

                                                                     pre-Exfoliation; extraction restriction

    b. [$_{CP}$ [C$_{[WH, EPP]}$ [$_{InflP}$ [Infl [$_{vP}$ EA$_{[WH]}$ [v [$_{VP}$ [V PP]]]]]]]]

                                                                          Exfoliation applies as Last Resort

    c. [$_{CP}$ EA$_{[WH]}$ [C$_{[\sout{WH, EPP}]}$ [$_{InflP}$ [Infl [$_{vP}$ ~~EA~~$_{[WH]}$ [v [$_{VP}$ [V PP]]]]]]]]

                                                                             surface result

Note that we make a prediction for cases where both the EA and the IA are local persons. Since the IA is encased in a PP, we can assume that the IA bearing the [PART] feature is licensed by the P^0 (the relational noun in Mayan-specific terms). I assume that the Set A marker on the RN is a doubled clitic that is prefigured by an Agree relation between the RN and the IA. As a result of the IA being licensed by the RN, we predict that the EA can be licensed by Infl^0 after Exfoliation, and that no crash arising from a PLC violation will occur.

This prediction is correct. In contrast to AF, which is banned if there are two local person arguments, AP$_{OBL}$ can be used in such a configuration:

(169) *Two local persons allowed with the AP$_{OBL}$*

    a. *Xa   xe   rat x-a-/i-to'-o   rïn.
       EMPH only 2S COM-B2S-/A1S-help-AF 1S
       Intended: 'Only you helped me.'

    b. Xa   xe   rat x-a-to'-o   w-ichin *pro*$_{1SG}$.
       EMPH only 2S COM-B2S-help-APO A1S-RN
       'Only you helped me.'   (adapted from Heaton 2017: 344)

A second prediction of the proposed analysis is that AF and the AP$_{OBL}$ have the same morphological exponence. This is correct. Recall that we observed that the suffixes that

appear in the AF and AP$_{OBL}$ are identical, sharing the same distribution depending on the properties of the stem of affixation. A unified analysis straightforwardly explains this: the Vocabulary Insertion rule of AF and the AP$_{OBL}$ is the same. In both constructions, Voice^0 is removed and a morpheme appears on the stem that is the elsewhere insertion of *v* (see (106)-(111), in section 2.3.3).

Returning to our mismatches, we now see why it is unproblematic for the proposed account that AP$_{OBL}$ can mismatch with active voice under sluicing, as shown at the beginning of this section. The syntax of AP$_{OBL}$ is identical to AF, so, much like the AF-Active mismatches discussed in section 2.2.3.3, the mismatch complies with featural non-distinctness.

Let us return now to the mismatches we discussed in section 2.2.3, where it was assumed that *AF* was present in the sluice, given the EA *wh*-remnant. Now that we have discussed the AP$_{OBL}$ and its syntax, we can delve more closely into the nature of these mismatches. I will illustrate this with the cases in which there is passive voice in the antecedent.

Consider the examples below. Since there is an EA *wh*-remnant in the target clause, we have to ensure that *either AF or AP$_{OBL}$* is in the silence. At first glance, there is no way to ensure via the manipulation of the remnant that the sluice contains AF rather than AP$_{OBL}$ or vice-versa (170)d:

(170)   *Passive—AF/AP$_{OBL}$ mismatches*
      a. Ri aq x-Ø-kam-is-äx   r-uma jun winäq. Ta-b'ij   pe chwe
         DET pig COM-B3S-die-CAUS-PAS A3S-RN a   person IMP-say DIR PREP.A1S.RN
         achike winäq x-Ø-kam-sa-n
         what   person COM-B3S-die-CAUS-AF
         'The pig was killed by some person. Tell me which person killed it!'
                                      non-sluice, AF in target clause

b. Ri aq x-Ø-kam-is-äx r-uma jun winäq. Ta-b'ij pe chwe
   DET pig COM-B3S-die-CAUS-PAS A3S-RN a person IMP-say DIR PREP.A1S.RN
   achike winäq x-Ø-kam-sa-n r-ichin!
   what person COM-B3S-die-CAUS-APO A3S-RN
   'The pig was killed by some person. Tell me which person killed it!'

non-sluice, AP$_{OBL}$ in target clause

c. *Ri aq x-Ø-kam-is-äx r-uma jun winäq. Ta-b'ij pe chwe
   DET pig COM-B3S-die-CAUS-PAS A3S-RN a person IMP-say DIR PREP.A1S.RN
   achike winäq x-Ø-u-kam-sa-j.
   what person COM-B3S-A3S-die-CAUS-ACT
   Intended: 'The pig was killed by some person. Tell me which person killed it!'

*non-sluice, active in target clause

d. Ri aq x-Ø-kam-is-äx r-uma jun winäq. Ta-b'ij pe chwe
   DET pig COM-B3S-die-CAUS-PAS A3S-RN a person IMP-say DIR PREP.A1S.RN
   achike winäq < x-Ø-kam-sa-n // x-Ø-kam-sa-n r-ichin>!
   what person COM-B3S-die-CAUS-AF COM-B3S-die-CAUS-APO A3S-RN
   'The pig was killed by some person. Tell me which person!'

sluice, Passive-AF/AP$_{OBL}$

There are two analytical reasons why this is not a problem. First, what we *can* conclude is that examples like the above involve a mismatch between an antecedent containing Voice^0 and an ellipsis site lacking Voice^0 (either AF or AP$_{OBL}$); i.e., either the Passive mismatches with AF, or it mismatches with AP$_{OBL}$. In either case, the mismatch complies with the proposed identity condition predicated on featural non-distinctness.

Alternatively, we could hypothesize that the sluices in these examples *must* be AF, because AP$_{OBL}$ in the sluice would necessitate there being a relational noun present in the ellipsis site that is not present in the antecedent. This kind of configuration is ruled out as shown by evidence from languages that allow P-stranding in general. Crucially in these languages, P-stranding under *sprouting* is impossible (Chung 2006; see Rosen 1976 example 15 for the first report of these data):

(171) *Chung's generalization—there is no P-stranding under sprouting*
   a. Jaida danced last night, but I don't know [with who]$_1$ <Jaida danced t$_1$>.
   b. *Jaida danced last night, but I don't know who$_1$ <Jaida danced [**with** t$_1$]>.
   c. Trixie is jealous, but no one knows [of who]$_1$ <Trixie is jealous t$_1$>.

d. *Trixie is jealous, but no one knows who$_1$ <Trixie is jealous [**of** t$_1$]>.

Having a Passive-AP$_{OBL}$ mismatch would be similar, since there would be a prepositional element (specifically, a relational noun) in the ellipsis site that is not present in the antecedent. Chung (2006) proposes a lexical requirement on ellipsis to rule this type of examples out (see van Craenenbroeck & Merchant 2013 for discussion and chapter 5 here). We cannot independently test examples like the above in Kaqchikel (given that P-stranding is impossible in general in the language), but Chung's generalization about P-stranding and sprouting seems robust. As a result, we could rule out the existence of Active-AP$_{OBL}$ and Passive-AP$_{OBL}$ mismatches and conclude that whenever there is an EA *wh*-remnant, we have in fact forced the use of AF and only AF in the ellipsis site.

One final note is necessary regarding the AP$_{OBL}$. One might wonder why it is illicit to have an IA encased in a RN with *active* Voice. In other words, what goes wrong in examples like the following?

(172) *Active voice, PP internal argument*
    *Voice$_{ACT}$ [PP IA] EA

The problem with this configuration appears to be related to the following issue—notice below that in an active voice configuration, it is not possible to have the thematic agent as a by-phrase (for discussion of a version of this problem, see Lasnik 1988):

(173) *No by-phrase in active voice*
    a. ***By Bianca del Rio** knocked on the door.
    b. *Someone knocked on the door **by Bianca del Rio**

I suspect that these two voice-related phenomena are related, but leave a more in-depth discussion for the future.

To summarize this subsection, we began by delving deeper into the syntax of AP_{OBL}, noting first that this voice can also mismatch with Active voice under sluicing. While initially this appeared problematic for the proposed analysis in this chapter, we established that the syntactic distribution of the AP_{OBL} is *identical* to AF. Building on this, I proposed that the syntax of AF and the AP_{OBL} is the same. Thus, mismatches involving the AP_{OBL} also comply with featural non-distinctness under sluicing.

2.5.2 All other possible voice combinations in Kaqchikel

In section 2.2.3, I noted that there are 20 conceivable voice combinations that one could construct under sluicing in Kaqchikel. There are three points that deserve discussion now, since we are at a juncture where we can discuss all these combinations: (i) mismatches that cannot be tested due to reasons that are internal to Kaqchikel grammar; (ii) mismatches that cannot be forced via the *wh*-remnant, and (iii) empirical gaps.

2.5.2.1 Mismatches that cannot be constructed at all

There are two important voice mismatches that we *cannot* test in Kaqchikel. As we observed at the beginning of this chapter, the impossibility of Active-Passive and Passive-Active mismatches under sluicing cross-linguistically has been crucial in establishing that syntax needs to be a part of the identity condition on ellipsis. However, neither of these can be constructed in Kaqchikel.

The way to test for the (un)availability of Active-Passive mismatches is by having an antecedent in the active voice and forcing passive voice in the ellipsis site via the use of a *wh*-remnant that corresponds to the agent—e.g., 'by who/ who by' in English. As I noted in footnote 59, however, I have been unable to elicit a *wh*-question targeting the agent of a passive from any of my consultants. Observe below that no *wh*-word coupled with the

relational noun that introduces the agent of a passive can be interpreted as a question targeting the agent. Instead, these strings are licit but are interpreted as reason or benefactive adjunct questions:

(174) *There is no 'by who'/'who by' question in Patzún Kaqchikel*

    a. Ri   aq x-Ø-kam-is-äx   r-uma   jun ixöq!
      DET pig COM-B3S-die-CAUS-PAS A3S-RN a   woman
      'The pig was killed by a woman!'

    b. Achike r-uma   x-kam-is-äx   ri   aq?
      WH   A3S-RN COM-B3S-die-CAUS-PAS DET pig
      'Why was the pig killed?' // # 'Who was the pig killed by?'

                                                             reason adjunct interpretation

    c. Achoq r-uma   x-kam-is-äx   ri   aq?
      WH   A3S-RN COM-B3S-die-CAUS-PAS DET pig
      'Who was the pig killed for?' // # 'Who was the pig killed by?'

                                                                benefactive interpretation

It is therefore impossible to test whether an Active-Passive mismatch is allowed, because the relevant strings are interpretable, but do not mean something akin to the target meanings of the (impossible) English '*Someone betrayed Robin, but we don't know by who.*' Instead, my consultants interpret sluices of this type as a reason adjunct sprout (174)b or a benefactive sprout (174)c. The impossibility of testing this mismatch is a mystery at the present time, but it is worth noting that it does not detract from the evidence from, and analysis of, those mismatches that *are* testable in Patzún Kaqchikel.

    I have been unable to find any source detailing how to construct the relevant *wh*-question (for example, no discussion exists in García Matzar & Rodríguez Guaján 1997 or Patal Majzul et al. 2000). Interestingly, Judith Aissen (p.c.) notes that she has also been unable to elicit the relevant kind of questions for Tz'utujil, which is closely related to Kaqchikel. One could imagine that this type of question—where the *wh*-phrase is the agent of a passive—would be a possible strategy to circumvent the extraction restriction on

external arguments of transitives. To my knowledge, however, no one has ever described this as a circumvention strategy in any discussion of the extraction restriction in Mayan. To the extent that this gap is substantial, as opposed to accidental—i.e., none of the syntactically ergative Mayan languages allow for the constructions of the relevant question—we need to investigate further what the source of the ban is. I leave this issue for future work.

Similarly, we cannot form Passive-Active mismatches in Kaqchikel for grammar internal reasons. The way to force these mismatches in languages like Spanish is by having passive voice in the antecedent and using a bare *wh*-remnant in the target clause that corresponds to the external argument of an active transitive. However, since Kaqchikel requires the use of Agent Focus or the oblique antipassive in such a configuration, there is no way to test the availability of Passive-Active mismatches, either.

To summarize, the two voice mismatches (Active-Passive; Passive-Active) that have formed the backbone of discussions about syntactic identity under sluicing cannot be tested in Kaqchikel. We could not conclude from this, however, that Kaqchikel allows voice mismatches freely. Recall that AP-Active mismatches—as opposed to Active-AF, AF/AP$_{OBL}$-Active, and Passive-AF mismatches—are ill-formed in the language, showing that we do find that syntax plays a role in regulating the availability of ellipsis. Clashes between Voice$_{AP}$ and Voice$_{ACT}$ are disallowed, while voice mismatches satisfying featural non-distinctness are permitted.

2.5.2.2 Mismatches that cannot be forced

A second issue worth discussing is that there are certain configurations in which it is not possible to syntactically force the relevant voice in the ellipsis site. The data that are

relevant here involve any mismatches where the AP voice would be used in the ellipsis site.

Recall from before that AP-Active mismatches are disallowed in Kaqchikel. It would be natural to ask, then, whether the AP voice can be inside the ellipsis site, as opposed to the antecedent, and mismatch with other voices—we would predict some of these combinations to be allowed, and others disallowed, depending on whether featural non-distinctness is satisfied. These possibilities are shown in the schemas below:

(175)   *AP voice in the ellipsis site (not tested)*
      a. … Voice$_{ACT}$ …   *wh*-remnant < … Voice$_{AP}$ … >   *prediction*: *
      b. … Voice$_{PASS}$ …   *wh*-remnant < … Voice$_{AP}$ … >   *prediction*: *
      c. … Ø (AF) …   *wh*-remnant < … Voice$_{AP}$ … >   *prediction*: ✓
      d. … Ø (AP$_{OBL}$) … *wh*-remnant < … Voice$_{AP}$ … >   *prediction*: ✓

Unfortunately, there is no *wh*-remnant that forces the use of the AP voice. In other words, there is no way to ensure via syntactic manipulations that the AP voice be inside the relevant ellipsis sites. I therefore have not attempted to test such examples.

One might wonder whether one could elicit the relevant data and ask consultants if the relevant *interpretations* are available—i.e., whether the sluices can have the meaning that is associated with the antipassive voice. Since these judgments would be delicate, I do not think that any conclusions drawn from such an exercise would be reliable description-wise, let alone in order to make conclusions about the identity condition on ellipsis. As we will observe in chapter 3, we must exercise extreme care when it comes to data where the remnant has not *forced* a specific syntactic content in the ellipsis site.

2.5.2.3 Empirical gaps

There are some empirical gaps in our overall picture of voice mismatches in Kaqchikel sluicing. In particular, I do not have data showing whether AP$_{OBL}$-AP$_{OBL}$/AF *matches* are

allowable under sluicing. Furthermore, I do not have data showing that a Passive-Passive match is well formed; such a configuration could be constructed via adjunct sprouting (e.g., 'Those flowers were bought by Ana, but I don't know where <they were bought by Ana>'). It is expected that these example types be well-formed, but I have not elicited the relevant data yet.

2.5.2.4 Summary: all voice mismatches in Kaqchikel

The table below specifies the possible and impossible voice combinations in Kaqchikel sluicing, along with (i) which mismatches *cannot be constructed* due to the specific nature of Kaqchikel grammar (i.e. Active-Passive), (ii) the voice combinations that *cannot be forced* via the *wh*-remnant, and (iii) empirical gaps. Notice that the ✓/* when an AP$_{OBL}$ voice is in the sluice is meant to signify that such a configuration would be a counterexample to Chung's generalization (see (170)-(171) and the surrounding discussion):

(176) *Sluicing in Kaqchikel—all voice combinations*

	Voice	Active	Passive	AF	AP$_{OBL}$	AP
ANTECEDEN	Active	✓	can't construct	✓	✓/*	can't force
	Passive	can't construct	no data	✓	✓/*	can't force
	AF	✓	can't construct	✓	✓/*	can't force
	AP$_{OBL}$	✓	can't construct	no data	no data	can't force
	AP	*	can't construct	can't force	can't force	can't force

(The top spanning header above Active/Passive/AF/AP$_{OBL}$/AP reads SLUICE.)

The table above shows that there remains some empirical work to be done. I plan to undertake such work in the future.

2.5.3 Do specifiers survive Exfoliation?

There is one important question concerning the Structural Change involved in Exfoliation which I have abstracted away from so far. The question is as follows: Does Exfoliation target the entire phasal projection, deleting in tandem its specifiers (as assumed here; Pesetsky 2017) or does Exfoliation remove only the phasal head, allowing the specifier to reattach to a lower projection (Pesetsky 2019, 2021)? Pesetsky notes briefly that this issue deserves further scrutiny, because there is some reason to believe that specifiers survive Exfoliation.

Given the novelty of Exfoliation as a syntactic hypothesis, there remain many questions to be answered about its proper formulation. If it is correct that AF is the result of $Voice^0$ Exfoliation, this analysis lends weight to a version of Exfoliation where the specifier does *not* survive. Recall that the omnivorous nature of $Infl^0$ agreement in AF crucially depended on the deletion of the copy of the IA in SpecVoiceP, severing the clitic-doubling chain with the clitic in $Infl^0$, and causing the latter to probe again (this time yielding the omnivorous agreement pattern discussed before). I proposed that this deletion prompts $Infl^0$ to probe again, whereupon it is able to target either argument. If, however, we assume that the IA in SpecVoiceP survives, we might run into a technical problem: why isn't agreement with the IA the only thing that ever surfaces in AF?

A possible solution to this conundrum would be to assume that $Infl^0$ inherits its features from C^0 (Chomsky 2008). In other words, $Infl^0$ does not probe until C^0 is merged. If we assume as well that all operations occur simultaneously, then there is a conceivable derivation where $Infl^0$ probes only after Exfoliation applies. In such a case, the IA and EA would be specifiers of the same head vP, and thus both would be accessible to the $Infl^0$ probe.

To summarize, I take the precise nature of Exfoliation in the narrow syntax as a work in progress. As more empirical domains are accounted for via the application of Exfoliation, the issues just raised will become clearer. For the particular case of Kaqchikel voices and AF, though, Exfoliation that results in the removal of specifiers delivers desirable results.

2.6 Summary and conclusion

In this chapter, we have explored a novel set of Kaqchikel data showing that Voice can mismatch in sluicing—a clause in the AF voice can mismatch with active or passive clauses. In light of these data, I argued for the following: (i) the Voice layer in AF clauses has been derivationally removed via Exfoliation (Pesetsky 2017, 2019, 2021), and (ii), the syntactic component of the identity condition on ellipsis is satisfied by featural non-distinctness, as opposed to featural identity. We have also seen that AF is a proper subpart of the voice system of Kaqchikel. In other words, AF clauses are neither active nor passive, but a distinct type of clause (in which VoiceP has been removed), and thus a type of construction that belongs in the discussion of the behavior of voice under ellipsis (Merchant 2013a).

Empirically, we have shown that broadening the scope of the study of ellipsis to the right kind of languages allows us to discover generalizations that would otherwise have remained undetected. It had mostly been assumed in the literature that Voice mismatches are banned *altogether* under sluicing (modulo the Chamorro and Malagasy cases; see 5.1 for discussion). Kaqchikel shows that this is not the case. There is nothing special about Voice mismatches that renders them illicit in general under sluicing. Rather, mismatches *at any level* that violate featural non-distinctness are banned. We took Active-Passive and

Passive-Active mismatches to be disallowed in languages like English because they involve a clash between Voice$_{ACT}$ and Voice$_{PASS}$; in a similar manner, AP-Active mismatches in Kaqchikel involve a clash between Voice$_{AP}$ and Voice$_{ACT}$. Conversely, voice mismatches satisfying featural non-distinctness are allowed, a possibility instantiated by the Kaqchikel data we discussed where AF clauses lack VoiceP. In the future, sluicing should be explored in other Mayan languages that display similar voice alternations as Kaqchikel, in order to attempt to replicate the data here and lend further empirical support to our generalizations. For example, Ranero 2019a reports data in K'iche' that are identical to Kaqchikel. Observe below that an AF-Active mismatch is also well-formed in this language:

(177) *AF-Active mismatch in K'iche' sluicing is well-formed* (Ranero 2019a)
    A: Ri a Lu' x-Ø-kam-sa-n   le ek'!
       DET CLF Pedro COM-B3S-die-CAUS-AF DET chicken
       'PEDRO killed the chicken!'
    B: La sitzij? Jas r-uuk' x-Ø-u-kam-sa-j   /*x-Ø-kam-sa-n
       Q truth WH A3S-RN COM-B3S-A3S-die-CAUS-TRANS   COM-B3S-die-CAUS-AF
       wi?
       FP
       'Really? With what did he kill it?'
    B′: La sitzij? Jas r-uuk' <x-Ø-u-kam-sa-j   wi>?
       Q truth WH A3S-RN COM-B3S-A3S-die-CAUS-TRANS FP
       'Really? With what?'

To reiterate, the analytical results of this chapter form the backbone of the rest of the dissertation and have far reaching consequences for our understanding of the conditions that govern the availability of ellipsis. The particularities of the grammar of Kaqchikel have allowed us to probe, within the Voice layer, into the nature of possible mismatches under ellipsis. Specifically, we showed that a clause lacking Voice can mismatch with a clause specified for Voice, since such a configuration satisfies featural non-distinctness. As we will see in chapter 3 when we expand our empirical domain into TP/Mittelfeld-level

mismatches in languages like English and Spanish, the proposed identity condition appears to be on the right track. For example, in a manner parallel to the Kaqchikel data, mismatches at the T-level that satisfy featural non-distinctness are allowed in English (e.g., a mismatch between a tensed clause and an infinitival), whereas mismatches that violate featural non-distinctness are banned in Spanish (e.g., clashes between fully tensed antecedents and ellipsis sites). Looking further ahead to chapter 4, we will see that featural non-distinctness as an identity requirement will form one of the pillars of an explanatory analysis of the behavior of [GENDER] mismatches with different noun pairs in Spanish and other languages with grammatical gender.

Chapter 3: Mismatches above Voice

In chapter 2, I concluded that the cross-linguistic facts concerning well-formed and ill-formed voice mismatches could not be explained by a strict syntactic identity condition. As a result, I proposed the identity condition in (1), which can account for the empirical generalizations:

(1) *Syntactic identity in ellipsis (first version; to be modified in chapter 4)*
The antecedent and material properly contained within the ellipsis site must be featurally non-distinct.

This chapter extends the empirical coverage of chapter 2. I will assess mismatches in ellipsis where the locus of the mismatch in the clause is *above* the Voice level. In brief, I will conclude that (1) can provide a uniform account of the data analyzed here as well, explaining possible and impossible mismatches across several languages and elliptical constructions. From an explanatory standpoint, this is an appealing consequence.

Let us start by recapping the key empirical takeaways from chapter 2, which should be kept in mind as we move our discussion forward. We first established that voice and argument structure mismatches are disallowed in English sluicing. A few examples illustrating the generalization are shown below:

(2) *Voice mismatches are disallowed* in English (Chung et al. 2011)
a. The candidate was abducted, but we don't know who by / by who.
b. *Somebody abducted the candidate, but we don't know by who.
c. Somebody abducted the candidate, but we don't know by who he was abducted.

(3) *Argument structure mismatches are disallowed in English*
a. *The jug broke, but I don't know who <broke the jug>.   (Merchant 2005)
b. *They loaded something onto the truck, but I couldn't quite make out with what <they loaded the truck>.
c. *They loaded one of the trucks with hay, but I couldn't quite make out onto which truck <they loaded hay>.   (Rudin 2019; based on Merchant 2005)

Merchant (2013)a argued on the basis of data like (2)-(3) above that there must be a syntactic component to the identity condition. The reasoning is straightforward: if one (only) assumes a semantic identity condition (e.g., one based on mutual entailment, Merchant 2001), it is unclear why the relevant examples in (2)-(3) would be ill-formed.

However, while syntax must be a component of the identity condition, our discussion of Kaqchikel voice mismatches led us to conclude that *strict* syntactic identity cannot be the requirement that regulates the availability of ellipsis. We were led to conclude this by the observation that a subset of voice mismatches are well-formed in Kaqchikel sluicing, in contrast to languages like English. These empirical results are summarized below:

(4)   *Voice mismatches in Kaqchikel sluicing*
    a.   AP-Act   *
    b.   Act-AF   ✓
    c.   AF-Act   ✓
    d.   Pas-AF   ✓

Putting the picture together, I proposed that the identity condition in (1) delivers the correct results, since it correctly rules out ill-formed English examples like those in (2)-(3) and correctly rules in the well-formed Kaqchikel examples summarized in (4). Importantly, an appeal to language specific requirements or stipulations was not required. This avoidance of language-specific stipulations fit well with our goal of proposing a uniform identity condition that regulates the distribution of ellipsis across languages and ellipsis types (i.e., a uniform identity condition holds for sluicing, VP ellipsis, NP ellipsis, etc., see chapter 1 and section 3.2 below).

In this context, we are now ready to zoom out of the Voice domain in the clause and consider other mismatches that have played a role in discussions about possible and impossible identity violations in ellipsis. Consider data like those below, which Merchant

(2001) took as support for a semantic, as opposed to syntactic, identity condition. The relevant takeaway from data like (5) is that the content of the ellipsis site is not fully identical to the antecedent.

(5)   *Mismatches above Voice* (adapted from Merchant 2001: 22)
    a. I can't play quarterback: I don't even know how <to play quarterback>.
    b. Invest now! We'll tell you how <to invest>.
    c. Eat (something), if you can figure out what <to eat>.
    d. Two or three men were crying. Others couldn't remember how <to cry>.

A puzzle arises when we are faced with the English data in (5), which seem to show that mismatches are allowed in some cases, in contrast to our ill-formed and now familiar voice and argument structure mismatches. If one proposes that the unacceptable voice and argument structure mismatches demonstrate that syntax *must* play a role in the identity condition, how does one account for the well-formed status of the data in (5)?

We will evaluate data akin to (5) in this chapter. This evaluation will proceed by comparing the identity condition proposed in this dissertation with another identity condition that appears to derive the asymmetry in mismatch availability that I highlighted above (Rudin 2019). In a nutshell, Rudin 2019 proposes a strict syntactic identity condition, although that condition is proposed to only range over elements within the VoiceP/*v*P domain. By proposing this partitioning of the clause—i.e., certain structure is considered by the identity condition and certain structure is not—Rudin (2019) derives the asymmetry. I will assess (and ultimately reject) this identity condition as an explanation for the asymmetry between unacceptable voice and argument structure mismatches on the one hand and acceptable mismatches on the other where the locus of the mismatching feature is higher in the structure. The proposal in Rudin 2019 makes a prediction about possible mismatches beyond English; namely, the prediction is that that we should find that

mismatches above the VoiceP/*v*P layer are allowed wholesale. A sample of data primarily from Spanish will show that this prediction is incorrect.

All in all, I will make theoretical, methodological, and empirical contributions in this chapter. I will conclude that there exists no wholesale allowance for mismatches above VoiceP/*v*P. Once I have established that Rudin 2019's solution for the mismatch asymmetry is not tenable, I will reassess all of the data that form the backbone of the proposal in Rudin 2019 and situate these data in the context of our conclusions from chapter 2. We will go through each of the data points in turn and show how the proposed identity condition in (1) can account for their well-formedness. Along the way, I will make methodological observations about determining the content of ellipsis sites via the manipulation of the antecedent and remnant, highlighting how certain cases need to be discussed with particular caution (recall the methodological question in section 1.4). Throughout the discussion, I will also make an empirical contribution by introducing additional data that serve as useful comparison points and which provide converging evidence that the identity condition I am proposing is on the right track.

The chapter is structured as follows. In section 3.1, I lay out the identity condition proposed in Rudin 2019 and show how it derives the mismatch asymmetry discussed previously. Section 3.2 clarifies the predictions made by that account. I show that cross-linguistic data, primarily from Spanish, show that these predictions are incorrect. In contrast, the identity condition defended here accounts for the relevant examples. The Spanish data where a source is not cited represent my own judgements and those of three other native speakers of Guatemalan Spanish. Moving on to section 3.3, I show there that the identity condition proposed in this dissertation also rules in correctly all of the well-

formed examples that are discussed in Rudin 2019. In a similar manner to the previous section, I introduce additional data that reinforce our confidence in the present proposal; whenever I discuss English data where a source is not cited, the judgments given reflect the intuition of three native speakers I consulted. Section 3.4 briefly considers an alternative analysis for some of the data that were evaluated in the preceding section. In 3.5, I expand further the empirical coverage of the identity condition proposed here, providing an account of [PERSON] mismatches that relies on the proper-containment clause of the condition. Finally, section 3.6 concludes and sets the stage for the discussion of mismatches in the nominal domain analyzed in chapter 4.

3. 1 A mismatch asymmetry and Rudin 2019's proposal

In this subsection, I discuss an asymmetry in the availability of mismatches under ellipsis and a proposal put forward in Rudin 2019 for an identity condition that derives this asymmetry.

Rudin 2019 discusses an array of sluicing examples that seem to show the following: mismatches in which the mismatching feature is above the VoiceP/vP level appear to be freely available in English, in contrast to voice and argument structure mismatches. Observe the contrast between the availability of tense mismatches on the one hand (6) and voice mismatches on the other (7); we will focus on the specifics of the relevant data later on. To begin, representative examples of possible T(ense) mismatches are shown below:[70]

---

[70] Most of the data in Rudin 2019 come from the UC Santa Cruz sluicing corpus, which provides natural language tokens of the kind of sluicing data that has been central to discussions on ellipsis since the seminal Ross 1969a (see Anand et al. 2021 for more examples). It would be desirable, of course, to have a comparable

(6) *Tense mismatches* (Rudin 2019: 266; a,b from Merchant 2001)[71]

    a. Sally **cooks** every night; she learned how <**to cook**> from her father.

    b. The baseball player went public with his desire **to be** traded. He doesn't care where <he {**is, will be**} traded>.

    c. Your favorite plant **is** alive, but you can never be sure for how long <your favorite plant **will be** alive>.

In (6)a-c, there is a mismatch at the T^0 level between the antecedent and ellipsis sites. The crucial takeaway is that, even though there is a tense mismatch, the examples are well-formed. In contrast, recall that voice mismatches under sluicing appear to be impossible in English (Merchant 2001, 2013). Representative examples are shown below:

(7) *Voice^0 mismatches*

    a. Someone **betrayed** Robin, but we don't know who <**betrayed** Robin>.

                                       ✓$\text{Voice}_{\text{ACT}}$ -$\text{Voice}_{\text{ACT}}$

    b. *Someone **betrayed** Robin, but we don't know by who/who by <Robin **was betrayed**>.   *$\text{Voice}_{\text{ACT}}$ -$\text{Voice}_{\text{PASS}}$

    c. Robin **was betrayed**, but we don't know by who <Robin **was betrayed**>.

                                       ✓$\text{Voice}_{\text{PASS}}$ -$\text{Voice}_{\text{PASS}}$

    d. *Robin **was betrayed**, but we don't know who <**betrayed** Robin>.

                                       *$\text{Voice}_{\text{PASS}}$-$\text{Voice}_{\text{ACT}}$

The asymmetry we have just highlighted can also be shown via argument structure mismatches, such as those we discussed in chapter 2 (see (3)).

    The following identity condition is proposed in Rudin 2019 to account for the asymmetry in mismatch possibilities:

---

repository of Spanish data to present in this chapter, but we will rely on prior descriptions and native speaker intuitions, and hope that such a project is undertaken in the future.

[71] The data in (6)a-b are deemed finiteness mismatches in Rudin 2019, while (6)c is called a tense mismatch. I will refer to all of these data moving forward as T^0/tense.

(8)   *Syntactic Condition on Sluicing (final)* (Rudin 2019: 269; bolding mine)

Given a prospective ellipsis site E and its antecedent A, non-pronunciation of the phonological content associated with any head h ∈ E is licit if at least one of the following conditions holds

      a. h did not originate within E's **eventive core**
      b. h has a structure-matching correlate i ∈ A.

The aspect of the condition in (8) that that derives the asymmetry and will be central to our discussion is that only heads which originate within the *eventive core* enter into the calculation of identity (8)a. The eventive core is defined as follows:

(9)   *Eventive Core (final definition)* (Rudin 2019: 271)

The eventive core of a clause is its highest vP that is associated with an event introducing predicate.

Note that Rudin 2019's definition in (9) posits that material under v^0 constitutes the eventive core; Rudin does not adopt proposals that split the topmost layers of the verbal domain into Voice^0 and v^0 (see Harley 2017). In commenting on the impossibility of voice mismatches, the paper states "*However, the most general way in which this account explains the impossibility of voice mismatches under sluicing is in terms of **mismatching v's** [emphasis mine]. A passive v cannot serve as a correlate to an active v, as they are not lexically identical, and so we predict the impossibility of voice mismatches under sluicing crosslinguistically regardless of the morphosyntactic reflexes of the passive in any particular language*" (Rudin 2019: 261). Since I am adopting proposals where Voice^0 is merged above v^0, I will consider in my evaluation that the eventive core includes Voice^0. Note that I am not changing Rudin's proposal in any significant manner—under a split Voice^0/v^0 analysis, Rudin would need to assume that Voice^0 is within the eventive core as well. Otherwise, the proposal would predict that voice mismatches are allowed freely under sluicing, contrary to fact.

Let us move on to the definition of a structure-matching correlate, relevant for part (8)b in the identity condition:

(10)   *Correlate* (Rudin 2019: 264)
        A node n can be a correlate for a head h iff at least one of the following conditions holds:
                a. n is a head and n and h are tokens of the same lexical item
                b. n is coindexed with h

We will focus on the proposal that the eventive core is privileged, as encapsulated in (11):[72]

(11)   *The eventive core (i.e. VoiceP/vP) is privileged*
        In calculating the satisfaction of the identity condition in ellipsis, the eventive core is privileged; material outside of the eventive core can mismatch without causing a violation of the syntactic identity condition.

Let us show how the application of the condition in (8) rules out the ill-formed voice mismatches. In the example below, there is a Voice$_{PASS}$ in the ellipsis site. This head is part of the eventive core, so it must find a structure-matching correlate in the antecedent (see (8)b). However, there is no structure-matching correlate in the antecedent, since the antecedent contains a Voice$_{ACT}$, not a Voice$_{PASS}$:

---

[72] Note that Rudin is not the first to claim that the eventive core (however defined) is special in the calculation of identity—the condition proposed in Chung 2013a already privileges the portion of the clause relevant to argument structure alternations (Chung 2013a's *Argument Structure Condition*; see chapter 5). More recently, den Dikken 2020 proposes an identity condition that states that "*ellipsis is grammatical only if all constituents of the predication structure involved in the ellipsis can find a match in the predication structure of the antecedent*". The arguments I will present against Rudin 2019's privileging of the eventive core transfer straightforwardly to any identity condition—e.g., Chung 2013a's and den Dikken 2020's—that seeks to carve out the eventive core / predicative core as the only domain that is relevant for the calculation of identity. See also my discussion in section 3.6 of Anand et al. 2021.

(12) *Deriving unacceptable mismatches through Rudin 2019's proposal*
    a. *Someone **betrayed** Robin, but we don't know by who/who by <Robin **was betrayed**>.
    b. *Antecedent*: Voice$_{ACT}$
    c. *Ellipsis site*: Voice$_{PASS}$ (*no structure-matching correlate in antecedent*)

In contrast, the mismatches in tense (6) do not involve elements in the ellipsis site within the eventive core. As a result, there is no requirement for the mismatching elements to have structure-matching correlates in the antecedent and the examples are thus ruled in.

To summarize, the analysis in Rudin 2019 capitalizes on one salient difference between the examples in (6) versus (7) to capture their distinct patterning. In both cases, there exists a mismatch in the content of the ellipsis site relative to the antecedent clause. However, the locus of the mismatches is at different levels of structure. Whereas the licit mismatch in (6) occurs at the T^0 level, the mismatch in (7) occurs at the Voice^0 level. Were we to assume that syntactic identity is calculated over the *entire* ellipsis site, we would expect that these types of mismatches should both be ill-formed, contrary to fact. The proposal in Rudin 2019 is that structure *up to a certain structural layer* is relevant in calculating identity under ellipsis: mismatches up to the eventive core (VoiceP/*v*P) level are intolerable, whereas mismatches above this domain are possible:

(13) *On privileging the eventive core*
    "The fact that some aspects of the interpretation of the ellipsis site, such as the verb and its arguments, are fixed, but other aspects, like modality, are slippery and indeterminate, suggests that the explanation of these left-peripheral mismatches is that identity requirements on sluices apply only to some subset of the elements in the elided TP. Specifically, I claim that identity conditions on ellipsis apply only to elements that originate inside what I call the *eventive core* of the elided clause— roughly speaking, the verb and its arguments." (italics in original; Rudin 2019: 267)

However, when we look beyond the particular English examples used in Rudin 2019, we find that some mismatches above the eventive core *are* ill-formed, contra the predictions

made by the above proposal. I will show that mismatches are uniformly ruled out when they instantiate a full featural *clash*, thus violating the identity condition proposed initially in chapter 2. The reason why the English data in (6) are licit is *not* because material outside of the eventive core can mismatch freely. Rather, the data in (6) are simply not the *kind* of mismatch that is banned in ellipsis; recall, for example, that we concluded in chapter 2 that mismatches are allowed when there is a head bearing a feature in the antecedent or ellipsis site that *mismatches*, but does not *clash,* with another head (more on this below).

To summarize this subsection, I laid out how the proposal in Rudin 2019 derives the mismatch asymmetry that is the primary empirical concern of this chapter. I will now show that this proposal makes incorrect predictions in cases where the condition proposed here makes correct ones.

3.2 Featural clashes everywhere are disallowed; the eventive core is not special

If it were true that there is something deep about the eventive core's primacy in calculating syntactic identity under ellipsis, we would expect that the eventive core would be privileged by all grammars, not just by English grammars. The proposal in Rudin 2019 therefore makes a strong empirical prediction: if we look beyond the single language under consideration, material outside the eventive core should be able to mismatch freely in ellipsis.

In light of this, let us make a brief excursus into the virtues of pursuing a unified identity condition for ellipsis. Suppose that after analyzing ellipsis examples in a wide array of languages, we ended up proposing a descriptively adequate condition of the following nature:

(14)   *Syntactic identity under ellipsis (hypothetical)*

       a. In languages A, B, C, the ellipsis site and antecedent must be featurally identical.

       b. In languages D, E, F, the ellipsis site and antecedent must be featurally identical *up to XP*.

The condition in (14) is a parameterized condition (of some sort), positing that the asymmetries we find regarding well-formed and ill-formed identity violations in ellipsis stem from languages respecting either (14)a or (14)b. While a condition like (14) may describe the facts and capture the observed data, it amounts to little more than a restatement of the explanandum, which seems undesirable from an explanatory perspective. My reasoning is a poverty of the stimulus argument: learners are never exposed to negative evidence regarding the kind of mismatches that are disallowed in their language under ellipsis—e.g., they never observe that a passive sluice cannot be anteceded by an active clause and vice-versa. We would expect this logic to extend to other kinds of mismatches as well. Let us assume, therefore, that the proposal in Rudin 2019 similarly does not lend itself to any kind of variation. Under this view, we must assume that UG encodes the syntactic identity condition under ellipsis and that this condition privileges the eventive core full-stop.

Let us expand the empirical picture, then, and discuss ellipsis examples in Spanish where we force a mismatch above the eventive core via the manipulation of the remnant. These data will show that privileging the eventive core in the identity condition on ellipsis is on the wrong track. Granted, the eventive core might form a privileged unit in other realms outside of the identity condition on ellipsis. For example, Rudin (2019) delves into Langacker 1974's discussion of syntactic rules (e.g., raising, fronting, lowering) that target elements in the eventive core exclusively (Langacker's "objective content"). Langacker's

observations are taken as conceptual backing for the identity condition put forth in Rudin 2019. The argument appears to boil down to this: since only material inside the eventive core appears to be relevant for certain transformations, then it is natural that only material inside the eventive core matters for the calculation of identity in ellipsis. However, absent a well-founded and explicit proposal as to *why* the conditions regulating the rules discussed in Langacker should also be relevant for ellipsis, this argument is lacking. This suggests that there might be an alternative that accounts for the asymmetry in mismatch availability that forms the core data discussed in this chapter. This alternative is the identity condition defended in this dissertation.

We can now move on to specific data which will adjudicate between the current proposal and the proposal in Rudin 2019. To reiterate, the current proposal predicts that a featural clash at the T^0 level, for example, should be ill-formed (as opposed to a featural mismatch that respects non-distinctness), whereas Rudin 2019's condition predicts that such a clash should be possible (given that T^0 is outside of the eventive core). Let us begin by observing that temporal adverbs force the use of specific tenses in Spanish. A past oriented adverb like *ayer* 'yesterday' is compatible with the past tense (15)a, while future oriented adverbs like *mañana* 'tomorrow' are compatible with the present habitual or the future (15)c. Using the present habitual or future with a past-oriented adverb is therefore ill-formed (15)b; using the past with a future oriented adverb is also ill-formed (15)d; and so is using the past/future with an adverbial that is compatible only with the habitual present (15)f. The presence of specific temporal adverbs forces us, then, to use an appropriate tense (note that tense marking in Spanish, including the future used here, is synthetic, in contrast to the English future, which is formed via the use of a modal):

(15)   *Tense and temporal adverbs in Spanish*

      a. Ayer   comí   tortillas.
         yesterday eat.1S.PAST tortillas
         'Yesterday, I ate tortillas.'

      b. *Ayer   como/comeré   tortillas.
          yesterday eat.1S.PRES/eat.1S.FUT tortillas

      c. Mañana   como/comeré   tortillas.
         yesterday eat.1S.PRES/eat.1S.FUT tortillas
         'Tomorrow, I eat/will eat tortillas.'

      d. *Mañana comí   tortillas.
          tomorrow eat.1S.PAST tortillas

      e. En general, yo como   tortillas.
         in general 1S eat.1S.PRES tortillas
         'In general, I eat tortillas.'

      f. *En general, yo comí/comeré tortillas.
          in general 1S eat.1S.PRES   tortillas

The controlled manipulation of adverbs will be our tool, then, to force a specific tense specification in the ellipsis site. With this in mind, let us turn to sluicing in Spanish. The examples in (16) below provide a baseline. (16)a and (16)b are a non-ellipitical and sluicing pair showing that a tense *match* between antecedent and ellipsis site is, as expected, well-formed. (16)c is a non-elliptical version of the utterance involving the use of the light verb *hacer* 'to do' plus *lo*, a 3rd singular pronominal (see Saab 2008: 117-122 for arguments that *hacer + lo* does not involve ellipsis). The function of utterances employing *hacerlo* is similar to the function of English utterances employing *do so*. I provide this type of example in the crucial test examples (17)-(18) as an additional control, to show that there is no general pragmatic restriction on switching tenses between two clauses that might explain the ill-formedness of tense clashes (17)b and (18)b:[73]

---

[73] An anonymous reviewer for an earlier version of this material commented that the ellipsis example in (16)b (where tense matches between the antecedent and the ellipsis site) sounds more natural if the verb in the

(16)   *Spanish sluicing—Tense match*

Context: Two people are discussing who is in charge of repairing cell phones at a shop. One of the speakers tells the other who is in charge on that day and on the next day.

a. Hoy, Ana **repara**   los celulares,   pero mañana,   no sé   quién
    today Ana **repair.3S.PRES** the cell.phones but   tomorrow NEG know who
    repara   los celulares.
    repair.3S.PRES the cell.phones
    'Today, Ana repairs the cell phones. Tomorrow, I don't know who repairs the cell phones.'

b. Hoy, Ana **repara**   los celulares,   pero mañana,   no SÉ   quién
    today Ana **repair.3S.PRES** the cell.phones but   tomorrow NEG know who
    <**repara**   los celulares>.
    repair.3S.PRES the cell.phones
    'Today, Ana repairs the cell phones. Tomorrow, I don't know who does.'

c. Hoy, Ana **repara**   los celulares,   pero mañana,   no sé   quién **lo**
    today Ana **repair.3S.PRES** the cell.phones but   tomorrow NEG know who **3S**
    hace.
    do.3S.PRES
    'Today, Ana repairs the cell phones. Tomorrow, I don't know who does so.'

We will now modify the examples so that the presence of a temporal adverbial in the remnant of the target clause forces the use of the clashing tense inside the ellipsis site. Note that the intended interpretation should be pragmatically recoverable, since the adverb is outside of the ellipsis site and serves, thus, to specify the kind of interpretation that is intended by the speaker. However, these sentences are unacceptable. In a nutshell, when we force a tense *clash* between the antecedent and ellipsis site, sluicing is disallowed:

(17)   *Spanish sluicing—Tense clash disallowed*

Context: Two people are discussing who is in charge of repairing cell phones at a shop. One of the speakers tells the other who is in charge on that day and who was in charge the day before.

---

remnant of the target clause (*saber* 'to know' in this case) bears sentential stress. This sentential stress is indicated orthographically here with capital letters. My consultants and I do not find a difference in acceptability between the stressed and non-stressed versions.

a. Hoy, Ana **repara**   los celulares,   pero ayer,   no sé   quién
    today Ana **repair.3S.PRES** the cell.phones but   yesterday NEG know who
    reparó   los celulares.
    repair.3S.PAST the cell.phones
    'Today, Ana repairs the cell phones, but yesterday, I don't know who repaired
    the cell phones.'

b. *Hoy, Ana **repara**   los celulares,   pero ayer,   no SÉ   quién
     today Ana **repair.3S.PRES** the cell.phones but   yesterday NEG know who
    <**reparó**   los celulares.>
      repair.3S.PAST the cell.phones
    Intended: 'Today, Ana repairs the cell phones, but yesterday, I don't know who
            did.   *T$_{PRESENT}$ − T$_{PAST}$

c. Hoy, Ana **repara**   los celulares, pero ayer,   no sé   quién **lo**
    today Ana **repair.3S.PRES** the cell.phones but   yesterday NEG know who **3S**
    hizo.
    do.3S.PAST
    'Today, Ana repairs the watches, but yesterday, I don't know who did.'

The examples in (17)a and (17)c that do not involve ellipsis show that the unacceptability

of the tense clash under sluicing cannot be the result of some general pragmatic avoidance

for switching tenses across clauses. Furthermore, the only difference between the

acceptable sluicing in (16)b and the unacceptable sluicing in (17)b is the existence of a

tense clash between antecedent and ellipsis site. Finally, a putative alternative pragmatic

condition that sought to account for the ill-formed status of an example like (17)b by

appealing to a requirement banning separate events in the sluice and the antecedent would

fail to account for the *well-formed* status of (16)b. Notice that in that example, the events

in the antecedent and sluice are separate, and the agent in the target clause is distinct from

the agent in the antecedent as well. Nevertheless, ellipsis is possible.

The example in (18)b shows that the effect is symmetrical as well: a T head in the

antecedent bearing a [PAST] feature is not allowed to clash with a T head in the sluice

bearing a [PRESENT] feature:

(18)   *Spanish sluicing—Tense clash disallowed*

Context: Two people are discussing who is in charge of repairing cell phones at a shop. One of the speakers tells the other who was in charge of making the repairs on the day before, as well as who is in charge in general.

a.   Ayer,   Ana **reparó**   los celulares,   pero en general, no sé   quién
yesterday Ana **repair.3S.PAST** the cell.phones but   in general NEG know who
repara   los celulares.
repair.3S.PRES the cell.phones
'Yesterday, Ana repaired the cell phones, but in general, I don't know who repairs the cell phones.'

b.   *Ayer,   Ana **reparó**   los celulares, pero en general, no SÉ   quién
yesterday Ana **repair.3S.PAST** the cell.phones but   in general NEG know who
< repara   los celulares.>
repair.3S.PRES the cell.phones
Intended: 'Yesterday, Ana repaired the cell phones, but in general, I don't
know who does.'   *$T_{PAST} - T_{PRESENT}$

c.   Ayer,   Ana **reparó**   los celulares,   pero en general, no sé   quién
yesterday Ana **repair.3S.PAST** the cell.phones but   in general NEG know who
lo   hace.
3S do.3S.PRES
'Yesterday, Ana repaired the cell phones, but in general, I don't know who does so.'

In sum, I assume that the feature geometry for tense in Spanish is such that we are dealing with a featural clash between $T_{[PRESENT]}$ and $T_{[PAST]}$ (as opposed to an alternative feature geometry where one of these two tenses involves the absence of a tense feature). The identity condition defended in this dissertation can thus account for the unacceptability of (17)b and (18)b.

In contrast, these results from Spanish could not be accounted for by the identity condition in Rudin 2019: the relevant mismatching elements lie outside of the eventive core, so it is predicted that they be able to mismatch under ellipsis, *regardless* of the feature geometry adopted for tense. Finally, I emphasize once more that appealing to some sort of (additional) pragmatic condition governing the impossibility of (17)b and (18)b does not seem like a fruitful line of explanation here, since the adverbial remnant makes clear what

the speaker's intended interpretation for the tense inside the sluice must be. These data show, then, that upon broadening the empirical scope of our investigation, there exists counterevidence to the proposal in Rudin 2019 that structure outside the eventive core is ignored by the identity condition.

Notice that the translations to the examples just discussed show that we cannot construct parallel examples in English to the Spanish tense clashes, since the use of VP ellipsis (VPE) would be required. Importantly, VPE is needed here, regardless of whether we force a tense mismatch (19)d or not (19)b:[74]

(19)   *English is (sometimes) the wrong empirical source*
    a. Today, Ana **repairs** the cell phones. Tomorrow, I don't know who does/will <**repair** the cell phones>. *VPE*
    b. *Today, Ana **repairs** the cell phones. Tomorrow, I don't know who <**repairs** the cell phones>. *sluicing; T match*
    c. Today, Ana **repairs** the cell phones. Yesterday, I don't know who did <**repair** the cell phones>. *VPE*
    d. *Today, Ana **repairs** the cell phones. Yesterday, I don't know who <**repaired** the cell phones>. *sluicing; T clash*

These VPE examples are irrelevant to the discussion at hand, since the *do*-remnant bears the tense specification that would clash with the antecedent in a well-controlled test case, whereas the verb within the ellipsis is bare. VPE involving a remnant modal *will* is equally uninformative, since the verb inside the ellipsis site is also tenseless. The question of why

---

[74] Notice that the problem does not lie in English lacking the inventory of heads that is necessary to test the mismatches (i.e., different flavors of tensed T^0). In other words, the issue is not that Spanish possesses a rich inventory of tensed T^0s, whereas English lacks such an inventory. Rather, what seems to be impossible in English is the possibility of constructing the controlled environment in which we could force a mismatch. In other words, English sluicing does not provide a window into the availability, or lack thereof, of tense clashes under ellipsis, because sluicing is disfavored in favor of VP ellipsis in these cases.

it is the case that VPE is needed in these English examples is interesting in its own right, but ultimately orthogonal to the main point here.

The contrast between Spanish and English, however, should drive home the point that the grammar of English precludes us (in this instance) of testing the outcome of featural clashes at the T^0 level. Spanish, however, provides the appropriate testing ground, and the examples in (17)-(18) show that the prediction made by Rudin's (2019) account is incorrect. The data favor the current proposal instead.

The Spanish facts illustrated previously are not merely a quirk of *sluicing*. Parallel conclusions can be drawn from other types of ellipsis where the remnant in the target clause is an element other than a *wh*-word. In what follows, I will consider TP-ellipsis where the remnant is an element like *también* 'also' or a polarity marker *sí/no* 'yes/no'. Before delving into the data, let us establish that an analysis of the sort shown in (20), below, is sound for the types of elliptical constructions that we will evaluate shortly (see Depiante 2004, Saab 2008, 2010b, 2016, Brucart & MacDonald 2012).[75] Here, the remnant is a clitic-

---

[75] There is evidence that all the Spanish TP-ellipsis examples we discuss involve structure in the ellipsis site, as opposed to a null pronominal—i.e., they are instances of surface anaphora in the sense of Hankamer & Sag 1976, not deep anaphora (see chapter 1 of this dissertation). First, an anaphor can take its antecedent from within the ellipsis site (the Missing Antecedent Phenomenon; Grinder & Postal 1971, Hankamer & Sag 1976; though see Merchant 2013b for caveats with this diagnostic):

(i)   a. *Sebas no compró ningún libro₁. Dice que *pro₁* está muy interesante.
           Sebas not bought any   book says that   is very interesting
           Intended: 'Sebas didn't buy any book. He says it's very interesting.'
        b. Sebas no compró ningún libro, pero yo sí compré un libro₁. *pro₁* Está muy
           Sebas not bought any   book but I yes bought a book   is   very
           interesante.
           interesting
           'Sebas didn't buy any book, but I did buy a book. It's very interesting.'
        c. Sebas no compró ningún libro, pero yo sí   ~~compré un libro₁~~. *pro₁* Está muy
           Sebas not bought any   book but I yes   is   very

left-dislocated element in the left periphery. The E-feature on the polarity head Σ (Laka 1990) licenses ellipsis in the syntax (Merchant 2001; see chapter 1):

(20)   *Spanish TP-Ellipsis*
       [$_{TopP}$ remnant$_i$ Top [$_{ΣP}$ Σ$_{[E]}$ [~~TP … Cl$_i$ +T~~…]]]

With this sketch of the structure in place, let us now turn to the relevant mismatches. Example (21) below involves TP-ellipsis where *también* 'also' is the remnant. We observe the same pattern as with sluicing: if we force a tense within the ellipsis that clashes with the tense in the antecedent, the example is ill-formed:

(21)   *Spanish TP-Ellipsis—Tense clash disallowed*
       a. Ayer   María **comió**   en el parque y anteayer   también
          yesterday María   **eat.3S.PAST** in the park   and day.before.yesterday also
          comió   en el parque.
          eat.3S.PAST in the park
          'Yesterday, Maria ate in the park, and the day before yesterday, she ate in the
           park as well.'
       b. Ayer María **comió** en el parque y anteayer también <**comió** en el parque>.
          'Yesterday, María ate in the park, and the day before yesterday, she did as well.

---

interesante.
interesting
'Sebas didn't buy any book, but I did. It's very interesting.'

Second, A'-extraction is possible from the ellipsis site—see chapter 4 on this diagnostic as applied to NP ellipsis:

(ii)   a. ¿Cuáles libros no compraste y   cuáles$_1$ sí compraste t$_1$?
            which books not bought   and which yes bought
            'Which books didn't you buy and which did you buy?'
         b. ¿Cuáles libros no compraste y   cuáles$_1$ sí <~~compraste~~ t$_1$>?
            which books not bought   and which yes

The possibility of extraction from the gap is strong evidence that there is structure in the gap. Additionally, the ellipsis types discussed here require syntactic control—i.e., they are surface anaphora in the sense of Hankamer & Sag 1976 (see Saab 2008). While this *in itself* is not a diagnostic for structure, it shows that these examples form a natural class with other ellipsis types—e.g. sluicing—that have been argued to involve structure in the ellipsis site.

c. Ayer,　　　María **comió**　　　en el　parque, y　mañana　　también **comerá**
　yesterday María **eat.3S.PAST** in the park　　and tomorrow also　　　**eat.1S.FUT**
en el　parque.
in　the park
'Yesterday, Maria ate in the park, and tomorrow, she will eat in the park as well.'

d. *Ayer, María **comió** en el parque, y mañana también <**comerá** en el parque>.
Intended: 'Yesterday, Maria ate in the park, and tomorrow, she will as well.'

$$*T_{PAST} - T_{FUTURE}$$

e. Ayer,　　　María **comió**　　　en el　parque, y　mañana　　también **lo**
yesterday María **eat.3S.PAST** in　the park　　and tomorrow also　　　**3S**
hará.
do.3S.FUT
'Yesterday, Maria ate in the park, and tomorrow, she will do so too.'

Similarly to our discussion of (17) through the prism of (19), notice that the translations

for the above examples show that we cannot construct parallel examples testing a featural

clash in English, since VPE is again required.

Next, observe the examples below, which illustrate the same point. The data in (33)

involve TP-ellipsis as well, but the remnant is a polarity marker *sí* 'yes':

(22)　　*Spanish TP-ellipsis—Tense clash disallowed*
　　　a. Mañana　　no　**iré**　　　　al　　parque, pero pasado mañana　sí　**iré**
　　　　tomorrow　NEG　**go.1S.FUT** to.the park　　but　after　tomorrow yes **go.1S.FUT**
　　　　al　　　parque.
　　　　to.the park
　　　　'Tomorrow, I won't go to the park, but the day after tomorrow, I will go to the
　　　　park.'

　　　b. Mañana no **iré** al parque, pero pasado mañana sí <**iré al parque**>.
　　　　'Tomorrow, I won't go to the park, but the day after tomorrow, I will.'

　　　c. Mañana　　no　**iré**　　　　al　　parque, pero ayer　　　　sí　**fui**
　　　　tomorrow　NEG　**go.1S.FUT** to.the park　　but　yesterday yes **go.1S.PST**
　　　　al　　　parque
　　　　to.the park
　　　　'Tomorrow, I won't go to the park, but yesterday I did go to the park.'

　　　d. *Mañana no **iré** al parque, pero ayer sí <**fui al parque**>.
　　　　Intended: 'Tomorrow, I won't go to the park, but yesterday, I did.

$$*T_{FUTURE} - T_{PAST}$$

e. Mañana   no   **iré**   al   parque, pero ayer   sí   **lo hice**
   tomorrow NEG **go.1S.FUT** to.the park   but   yesterday yes **3S do.1S.PAST**
   'Tomorrow, I won't go to the park, but yesterday I did go to the park.'

The takeaway from these examples is that elliptical constructions beyond sluicing show that full-featural clashes are ill-formed, even when the relevant features that clash are located above (and outside of) the eventive core. These examples show, then, that the solution proposed by Rudin 2019 for the asymmetry in mismatch availability at different structural levels is not tenable.

I am not the first to notice and discuss the unavailability of T^0-level mismatches in Spanish. Among others, Brucart (1987) and Murguia (2004) have made the exact same point before me. Murguia, for instance, shows that tense clashes are also disallowed in configurations involving complex tenses (i.e., auxiliary plus verb constructions). I have added the control example in (23)a to show that the non-elliptical version of (23)b is acceptable:

(23)   *Spanish TP-ellipsis: Tense mismatch disallowed; complex tense*
    a. En el pasado, María **ha**   leído mucho y   Elena en el   futuro
      in the past   María **have.PRES** read a.lot   and Elena in the future
      habrá   leído mucho también.
      have.FUT read a.lot   also
      'María has read a lot in the past, and in the future, Elena will have read a lot as well.'
    b. *En el pasado, María **ha**   leído mucho y   Elena en el   futuro
      in the past   María **have.PRES** read   a.lot   and Elena in the future
      <**habrá**   leído mucho> también.
      have.FUT read a.lot   also
            (adapted from Murguia 2004:86 *apud* Saab 2016)

Moving on, the results presented in Saab 2016 also cast doubt on the proposal in Rudin 2019. Saab argues against a semantic identity condition on ellipsis by showing that certain tense mismatches are impossible in Spanish TP-ellipsis *even* if the proposition in the antecedent *entails* the proposition in the target clause. The relevant examples involve a

clash between present and past tenses. The relevant T_PRESENT is the so-called *historical/narrative* present, which has a past interpretation and is used in specific discourse contexts. A baseline example showing the use of this particular tense is below:[76]

(24) *Historical present in Spanish* (adapted from Saab 2016: 372)
Adiviná qué me pasó ayer! Estoy tomando una
guess.IMP what to.me happened yesterday 1S.be.PRESENT drinking a

cervecita en el bar y entonces veo a mi mujer besándose
beer.DIM in the bar and then 1S.see.PRESENT A my wife kissing

con mi mejor amigo.
with my best friend
'Guess what happened to me yesterday! I am drinking a beer at the bar and then I see my wife kissing my best friend.'

The crucial observation is that using the historical present is impossible in the kind of interruption illustrated below. Speaker B must use the formal past in her interruption:

(25) *Constraints on the historical present in Spanish* (adapted from Saab 2016: 375)
A: Adiviná qué me pasó ayer! **Estoy** tomando una
  guess.IMP what to.me happened yesterday **1S.be.PRESENT** drinking a
  cervecita en el bar…
  beer.DIM in the bar
  'Guess what happened to me yesterday! I am drinking a beer at the bar…'

B: Qué casualidad! Ayer Juan también **estaba/#está**
  what coincidence yesterday Juan too **3S.be.PAST/3S.be.PRESENT**
  tomando una cervecita en el bar.
  drinking a beer.DIM at the bar
  'What a coincidence! Yesterday, Juan was also drinking a beer in the bar.'
  #'What a coincidence! Yesterday, Juan is also drinking a beer in the bar.'

We now have a relevant controlled environment in which to test whether TP-ellipsis is well- or ill-formed when a full-featural clash between the T^0 in the antecedent and the T^0 in the ellipsis site is forced. As is predicted by the approach to the identity condition

---

[76] While we are not speakers of Argentinean Spanish, my consultants and I agree with the judgments presented in Saab (2016).

170

proposed in this dissertation, such a mismatch is impossible. A historical T$_{\text{PRESENT}}$ cannot

serve as the antecedent for an ellipsis site that must contain a T$_{\text{PAST}}$, even though the

proposition in the antecedent clause containing the historical present entails the proposition

in the target clause containing the past (once the contribution of *también* 'too' is taken into

account):

(26)   *Historical T$_{\text{PRESENT}}$ - T$_{\text{PAST}}$ clash in Spanish* (adapted from Saab 2016: 377)

A: Adiviná   qué   me   pasó   ayer!   **Estoy**   tomando una
guess.IMP what to.me happened yesterday **1S.be.PRESENT** drinking   a
cervecita en el   bar…
beer.DIM in the bar
'Guess what happened to me yesterday! I am drinking a beer at the bar…'

B: Qué casualidad! Ayer   yo también <***estaba/#estoy***   tomando
what coincidence yesterday I   too   **1S.be.PAST/1S.be.PRESENT** drinking
una cervecita en el   bar.>
a   beer.DIM   at the bar   *historical T$_{\text{PRESENT}}$ − T$_{\text{PAST}}$
Intended: 'What a coincidence! Me too.'

The relevance of Saab (2016)'s examples should be clear: mismatches above the eventive

core are disallowed in Spanish, contra the predictions made by the proposal in Rudin 2019.

Overall, what is crucial in all cases is the *type* of mismatch—full clashes are ill-formed,

whereas we will observe that mismatches that satisfy featural non-distinctness are well-

formed (see discussion below in 3.3.1).

Further evidence for the impossibility of tense clashes in Spanish ellipsis is provided

in Pérez-Jiménez & Moreno-Quibén 2012. The authors of this work analyze *free exceptives*

and provide evidence that they involve (obligatory) clausal ellipsis (see section 3.3.5 below

for further discussion of exceptives in relation to polarity mismatches). The example in

(27)a is a baseline, showing that a temporal adverbial can be a remnant in the except-clause

(provided there is no clash in tense features; in this case, both the antecedent and the ellipsis

site contain the habitual present); this shows that temporal adverbials can be manipulated

in order to force a particular tense within the ellipsis site. As shown by (27)b, a clash

between $T_{PRESENT}$ and T_{PAST} is disallowed in this construction; the use of the adverbial

anteayer 'the day before yesterday' forces a past tense in the ellipsis site, leading to a

violation of the identity condition proposed here:

(27)   *Spanish exceptives disallow tense clashes*
      a. Los trabajadores comen aquí, excepto Juan los lunes   <no come aquí>.
        the workers   eat.3P here except   Juan the Monday NEG eat.3S here
        'The workers always eat here, except Juan on Mondays.'
      b. *Los trabajadores comen aquí, excepto Juan anteayer
        the workers   eat.3P here except   Juan day.before.yesterday
     <no   comió aquí>.
      NEG ate.3S here
      Intended: 'The workers always eat here, except Juan the day before yesterday.'
                    (adapted from Pérez-Jiménez & Moreno-Quibén 2012)[77]

Going over the evidence that exceptives involve ellipsis would take us too far afield.

However, an elliptical analysis of exceptives has been defended not only for Spanish, but

also English (Stockwell & Wong 2020; see Harris 1982, Merchant 2001: 107, fn. 12),

Malagasy (Potsdam & Polinsky 2019, Potsdam 2018, 2019) and Egyptian Arabic (Soltan

2016); see 3.3.4 below. If these proposals are on the right track and exceptives do involve

ellipsis, then the identity condition proposed in this dissertation makes the right predictions.

It correctly rules out ill-formed examples like (27)b; in contrast, the proposal in Rudin 2019

that only the eventive core enters into the calculation of identity in ellipsis incorrectly rules

---

[77] I have made two modifications to this example. First, I have made explicit the ellipsis site; note that I

include negation in the ellipsis site (see 3.3.4 for discussion of polarity mismatches in exceptives). Second, I

removed the adverbial 'always' from the antecedent, since it introduces an independent problem for the test

case in (27)b—the adverbial 'always' is incompatible with the remnant 'day before yesterday'. By removing

the adverbial, we can focus solely on the tense clash and construct a more minimal pair.

in ill-formed examples like (27)b, given that the locus of the mismatch is outside the eventive core.

Let us now summarize the results of this sub-section. We established that tense clashes in Spanish ellipses (sluicing, TP-ellipsis, and exceptives) are ill-formed. Given a specific conception of the feature-geometries under consideration (e.g., present and past tense are both featurally specified T^0 heads), this is correctly predicted by the identity condition proposed in this dissertation. The table below shows the T^0 mismatches we surveyed in this chapter. While I have not illustrated every conceivable tense clash in Spanish, the contrast with English will become even clearer when we delve into the English data in the next subsection:

(28)   *Tense mismatches in Spanish ellipsis*

Status	Antecedent	Ellipsis site	Language	Example
*	$T_{PRESENT}$	T_{PAST}	Spanish	(17)B', (27)b
*	T_{PAST}	$T_{PRESENT}$	Spanish	(18)B'
*	T_{PAST}	T_{FUTURE}	Spanish	(21)d
*	T_{FUTURE}	T_{PAST}	Spanish	(33)d ch. 1
*	historical $T_{PRESENT}$	T_{PAST}	Spanish	(26)

In sum, these empirical results from Spanish serve to adjudicate between the present proposal and the identity condition advocated in Rudin 2019 for the mismatch asymmetry that we laid out in 3.1. In few words, the asymmetry should not be derived via an identity condition that assigns a special status to the eventive core.

We are now in a position to return to the English examples that formed the empirical backbone of Rudin 2019 and show how they are accounted for by the identity condition proposed in this dissertation.

3.3 Back to English and some cross-linguistic detours

English corpus data form the empirical backbone of Rudin 2019. These well-formed sluicing examples display a range of mismatches in which the locus of the mismatch is above the eventive core. This fact leads the author to propose an identity condition that can capture the asymmetry in mismatch availability that we have been discussing (compare (3) and (5) with (6)). In this subsection, I will assess the data discussed in Rudin 2019 and reach two conclusions. First, I will conclude that the identity condition defended in this dissertation can account for the data provided in Rudin 2019. This conclusion, in combination with our results from chapter 2 and 3.2 above, provides us with solid footing to pursue broader empirical coverage in chapter 4, where we will zoom in on featural mismatches at the nominal level in [NUMBER] and [GENDER] that comply with featural non-distinctness. Second, I will conclude that some of the evidence provided in Rudin 2019 cannot, in fact, tease apart different formulations of identity conditions under ellipsis, given that English is frequently an ill-suited empirical testing ground to adjudicate between proposals.

At this juncture, readers might be concerned that the strategy that will be used in what follows is worryingly unrestrictive—could we not simply assume that every mismatch that is allowed is encoded via the presence/absence of a dedicated projection/feature, and mismatches that are not allowed reflect value conflicts among extant features, in an entirely post-hoc way? Recall, for example, my comments before that a specific feature geometry for Spanish tense must be adopted in order to attribute the ill-formed status of certain examples to a violation of the proposed identity condition—e.g., present and past tenses are both featurally specified values of the tense feature borne by T^0 and cannot mismatch under ellipsis, since such a configuration results in a featural clash. This concern is not

unjustified, but it does not undermine the current proposal altogether. For one thing, the Spanish data in 3.2 (and out Kaqchikel results, in particular the well-formed status of AF-Active mismatches) show that Rudin's proposal does not work when the empirical domain is broadened, so an alternative is required to begin with. If our specific alternative faces methodological challenges, this could only stand as a drawback relative to some *other* extant competitor that also explains Spanish and Kaqchikel (see chapter 2). More importantly, however, the maneuver of assuming presence/absence whenever a mismatch is allowed is not as freely-available as the foregoing discussion might suggest. For example, suppose we found that in Spanish, present and past clashes *were* allowed to mismatch. We don't think that an account appealing to, e.g., PresentP and PastP, each present and absent in a complementary set of clauses, would seem reasonable (compare this with the JussiveP-based approach to illocutionary mismatches, which enjoys greater plausibility; see 3.3.3 below). That is, it is not true that we have no priors regarding which projections are privative, and which stand in the conflicting-features relation, prior to and independent of their behavior under ellipsis.

With this in mind, the tolerated mismatches from Rudin 2019 that we will assess are the following: mismatches in tense/finiteness (3.3.1), modality (3.3.2), illocutionary force (3.3.3), and polarity (3.3.4). Where relevant, I will bring in cross-linguistic data that complement the discussion of the English examples, expanding thus an empirical picture that, in the end, will provide converging evidence for the non-distinctness approach to identity advanced in this dissertation. In section 3.3.5, I will briefly consider mood mismatches in Spanish, a type of data not discussed in Rudin 2019, and show how these

are fully consistent with our identity condition. Finally, the results will be summarized in 3.3.6.

3.3.1 Tense mismatches

Let us start by analyzing the nature of the mismatches underlying the data in (6) (the well-formed tense mismatches), in contrast with the nature of the mismatches underlying the data in (7) (the ill-formed voice mismatches).

According to the description in Rudin 2019 for the content of the relevant ellipsis sites, the tense mismatches in (6) (repeated as (29) below) involve a mismatch between tensed and tenseless clauses, or between a tensed clause and another clause where the verb is bare, but the presence of the modal *will* results in a future interpretation.[78] As we can see in (29), none of the English examples involve a *clash* between two *featurally specified* T^0 heads. Rather, the way to characterize these well-formed mismatches is that (i) a clause containing a tensed verb can mismatch with a clause containing a tenseless verb, (29)a and (29)b (when the sluice contains a tensed *is*), (ii) a clause containing a tenseless verb can mismatch with a clause containing a tenseless verb and an additional modal (29)b (when the sluice contains a modal *will* and a bare *be*), and (iii) a clause containing a tensed verb can mismatch with a clause containing a tenseless verb and an additional modal (29)c:

(29)　*Allowable mismatches at the T^0 level involve non-distinct T^0s (repeated from (6))*
　　　a. Sally **cooks** every night; she learned how <**to cook**> from her father.
　　　　　　　　　　　　　　　　　　　　　　　　　　　$T_{PRESENT} - T_{\emptyset}$
　　　b. The baseball player went public with his desire **to be** traded. He doesn't care
　　　　where <he {**is, will be**} traded>.　　　　　$T_{\emptyset} - T_{PRESENT} / T_{\emptyset} - T_{\emptyset} + modal$

---

[78] Note that the languages I am discussing here make absolute tense distinctions. In languages with relative tense systems, it might be difficult to force tense clashes under ellipsis.

c. Your favorite plant **is** alive, but you can never be sure for how long <your favorite plant **will be** alive>. $T_{PRESENT}$ - $T_{\emptyset}$ + modal

I will return to the content of the ellipsis sites below (see 3.3.2). For now, notice that in none of the examples above does there exist a tense mismatch between the antecedent clause and the ellipsis site in the following sense: (29)a-c do not involve *full featural clashes* between the T^0 in the antecedent and the T^0 in the ellipsis site, such that T^0 in the antecedent bears a feature that clashes with a feature borne by T^0 in the ellipsis site or vice-versa. I add the following example showing that an antecedent with a gerund can mismatch with an ellipsis site containing an infinitival:

(30)   I love **going** to the beach, I just don't know when <**to go** to the beach>. $T_{\emptyset}$ - $T_{\emptyset}$

We observe, then, that the nature of the English tense mismatches in the well-formed examples in (29)-(30) is fundamentally different from the nature of the voice and argument structure mismatches that are ruled out. In (29)-(30), there is no full featural clash. This fact makes these examples similar to the allowable voice mismatches in Kaqchikel, which I analyzed as instances where a clause lacking $Voice^0$ altogether mismatches with a clause containing a featurally-specified $Voice^0$.

In contrast, recall the voice mismatches in *English*: these examples were correctly ruled out by the identity condition defended here because $Voice^0$ in the antecedent bears a feature that clashes with a feature borne by $Voice^0$ in the ellipsis site and vice-versa. To be explicit once more, assume that $Voice_{ACT}$ and $Voice_{PASS}$ are heads bearing feature bundles—i.e., a set of features within one syntactic atom which contain distinct values of the same feature(s), resulting thus in the properties that differentiate active clauses from passive clauses. It should be clear, then, that the examples below instantiate a full featural clash. This full featural clash contrasts with the examples in (29), where no such clash exists:

(31)   *Voice mismatches involve a full featural clash*

   a. *Someone **betrayed** Robin, but we don't know by who/who by <Robin **was betrayed**>.   *Voice$_{ACTIVE}$ -Voice $_{PASSIVE}$

   b. *Robin **was betrayed**, but we don't know **who** <betrayed Robin>.
                                                     *Voice $_{PASSIVE}$ -Voice$_{ACTIVE}$

We therefore observe that there is an asymmetry between the *type* of mismatch in (29)-(30) (and the well-formed Kaqchikel voice mismatches) and the data in (31). While the proposal in Rudin 2019 attributes the difference in status among the relevant examples to a fact about the structural domain that enters into the calculation of syntactic identity in ellipsis, the identity condition defended here (repeated below) suffices:

(32)   *Syntactic identity in ellipsis (repeated from (1))*
   The antecedent and material properly contained within the ellipsis site must be featurally non-distinct.

Let us remind ourselves of the kind of examples that the identity condition in (32) rules in and the kind of examples it rules out. Ellipsis under (32) is allowed in several configurations: (i) when there is a mismatch between the presence and absence of a head bearing a feature bundle (e.g., the well-formed Kaqchikel voice mismatches); (ii) when both antecedent and ellipsis site have a specific head H, but one of these heads does not bear the relevant feature (e.g., the tense mismatches in English like (29)b); and (iii) when both antecedent and ellipsis have a specific head H, but neither bears the relevant feature (e.g., the English example in (30)). The well-formed data in (29)-(30) are thus correctly ruled in. Conversely, featurally distinct antecedents and ellipsis sites are disallowed given (32). In other words, ellipsis is impossible when there is a featural clash between a head bearing a feature bundle in an antecedent and the same head bearing a different feature bundle in an ellipsis site. The ill-formed data in (31) are thus correctly ruled out.

A question that arises at this juncture is whether cases of ellipsis besides sluicing could provide evidence that *English*, not just Spanish, also disallows featural clashes above the eventive core (recall our failed attempts at constructing the relevant sluicing examples in English in (19)). As discussed in Stockwell and Wong (2020), stripping (also called *bare-argument ellipsis*; see Hankamer & Sag 1976, Johnson 2019; see also Ortega Santos et al. 2014 for *wh*-stripping specifically) seems to provide that evidence. Consider the contrast between the examples below:

(33)   *Tense match in stripping*
        A: We **meet** in this building today.
        B: That's right! And where **do we meet** later today?
        B′: That's right! And where later today <**do we meet**>?

(34)   *Tense clash in stripping*
        A: We **met** in this building yesterday.
        B: That's right! And where **do we meet** later today?
        B′: *That's right! And where later today <**do we meet**>?   *T_{PAST} - $T_{PRESENT}$

The identity condition predicated on featural non-distinctness proposed here can account for the contrast between the well-formed examples in (33) involving a tense match and the ill-formed examples in (34) involving a tense clash. While a discussion of the precise derivation of stripping would take us too far afield, the contrast between the examples above shows that some ellipsis types in English also provide support for the analysis defended here. Since the relevant mismatches are outside of the eventive core, Rudin's (2019) proposal provides no purchase on why the contrast between (33) and (34) would hold.

To summarize this subsection, we assessed a collection of examples from Rudin 2019 involving tense mismatches. We concluded that the data can be accounted for by our non-distinctness approach to the identity condition. In the next subsection, we will assess

mismatches in modality. Similarly to our conclusions with the tense mismatches, we will observe that those data are accounted for by the identity condition we have defended thus far.

3.3.2 Modality mismatches

In this subsection, I will discuss mismatches in modality. Before delving into the specifics of the data, it is necessary to highlight a difference between the Spanish mismatches discussed in 3.2 and a subset of the English data that will concern us moving forward. In many of the English examples analyzed in Rudin 2019, it is impossible to know what the content of the sluice *must* be. Whereas we were able to construct controlled examples in Spanish, where we force a specific content for the ellipsis site via the manipulation of temporal adverbs or conversational turns, no such control is possible in a subset of the data I will assess here. As we will see, this limitation of the English mismatches will be important for our conclusion that these cases cannot adjudicate between different formulations of the identity condition. In what follows, I first present the examples and the content of the ellipsis sites as interpreted in Rudin 2019; if necessary, I provide additional, plausible interpretations.

Consider first cases in which there is no overt modal in the antecedent, but a modal interpretation is available for the sluice:

(35)   *Mismatch Type: Appearance of modality in the sluice* (Rudin 2019)
        Sally knows that there is always the potential for awful things **to happen**, but she doesn't know when <awful things {**will, might**} happen>.

We can account for this type of mismatch straightforwardly, since it instantiates another case of mismatches involving the presence versus absence of featural content. Whereas the antecedent has no ModP projection, the sluice does.

(36)   *Mismatch Type: Appearance of modality in the sluice (version 1)*
        Antecedent: $[_{XP} \ldots [YP]]$ *no Mod^0*
        Sluice: $[_{XP} [_{ModP} \ldots [YP]]]$ *Mod^0*

Notice that it would not matter whether we took a cartographic approach and assumed that all clauses project ModP—if the ModP in the antecedent is featurally empty, we would still not incur in a feature clash (see Cinque 1999 for an influential proposal for the placement of different modals in the clause).

The careful reader might wonder whether the presence of a modal in the sluice truly involves only the presence of a feature bundle heading ModP. For example, Chung (2006) argues from the unavailability of P-stranding under sprouting that the ellipsis site cannot contain lexical material that is not present in the antecedent (see van Craenenbroeck and Merchant 2013 for discussion; see chapter 5 for my analysis of Chung's generalization):

(37)   *No P-stranding under sprouting* (adapted from Chung 2006: 78)
        a. *Last night he was very afraid, but he couldn't tell us what <he was afraid **of**.>
        b. *We're donating our car, but it's unclear which organization <we are donating our car **to**.>
        c. *She phoned home, but they weren't sure which city <she phoned **from**.>

If Chung's lexical constraint is indeed correct and a nebulous modal such as the one interpreted inside the sluice in (35) should not be treated as a feature bundle, then this would constitute a challenge for the present approach.

Note, however, that alongside the possibilities for elided material given by Rudin (and reproduced in (35), above), the possibilities given in (38) below for the contents of the ellipsis sites seem entirely natural for the examples under consideration. If these are possible sources for the sluice, then we have a mismatch involving a tenseless antecedent and a tensed sluice, parallel to the tense mismatches that were discussed in 3.3.1:

(38) *Mismatch Type: Appearance of modality in the sluice* (*additional interpretations)*
    Sally knows that there is always the potential for awful things **to happen**, but she doesn't know when <awful things **Ø/do** happen>.

The point in providing the additional interpretations in (38) for the sluice under consideration is important—as long as there is at least one structure that could underlie the sluice that conforms to featural non-distinctness, *that* structure is sufficient to explain the well-formedness of the example. In other words, it is hard to construe a pragmatic state of affairs in which one of the proposed contents for the ellipsis site is appropriate and the other is not (i.e., these different alternatives seems to be interchangeable from a communicative standpoint). As a result, one could not assert that the content of the ellipsis site must be as in (35), as opposed to (38).

The examples characterized in Rudin 2019 as involving the appearance of modality fall neatly under the account proposed here, then, and need not be explained by attributing a privileged status to the eventive core. These examples can instead be analyzed as follows: there is a mismatch between the absence of finite T^0 in the antecedent and a $T_{PRESENT}$ in the sluice. There is no full featural clash and the example is correctly ruled in:

(39) *Mismatch Type: Appearance of modality in the sluice (version 2)*

| Antecedent: | $[_{TP} [T_\emptyset \dots [YP]]$ | *no finite T^0* |
| Sluice: | $[_{TP} [T_{PRESENT} \dots [YP]]$ | *finite T^0* |

The same point holds for the example of tense mismatch in (29)c, which we return to (and repeat) below as (40). In (40)a, I provide the characterization of the ellipsis site in Rudin 2019. I have consulted speakers who judge the non-ellipsis version with a matching tense as acceptable (40)b, albeit slightly strange, while others find it unnatural. The content of the ellipsis site could thus be as in (40)c:

(40)   *Allowable mismatches at the T^0 level ((40)a repeated from (29)c)*
    a. Your favorite plant **is** alive, but you can never be sure for how long <your favorite plant **will be** alive>.
    b. %Your favorite plant **is** alive, but you can never be sure for how long your favorite plant **is** alive.
    c. Your favorite plant **is** alive, but you can never be sure for how long <your favorite plant **is** alive.>

I hypothesize that the strangeness of (40)b is related to the predicate 'to be alive' and is orthogonal to ellipsis. Examples parallel to the above with a different predicate seem perfectly natural with the tense match:

(41)   *Allowable mismatches at the T^0 level*
    a. Sally **is** at home, but you can never be sure for how long she **is/will be** at home.
    b. Sally **is** at home, but you can never be sure for how long <she **is/will be** at home>.

The takeaway here should be clear. If there is a source for the sluice where there is no mismatch, then there is no problem to be explained, since we have no structural diagnostic that *forces* a mismatch. Note as well that the example in (40)a cannot be considered a controlled manipulation that ensures that there must be a modal in the sluice; i.e., the characterization in Rudin 2019 is a logically possible one, but there is nothing that ensures that it holds. Consider how this example is different from the Spanish data involving a mismatch between the historical present and past tenses (section 3.2). In these data, the semantics/pragmatics governing the use of the historical present are strong enough that there is *no way* to fill in the content of the ellipsis site with a featural match—in other words, a mismatch is forced. In contrast, I am assuming that the semantics/pragmatics governing the use of the predicate 'be alive' in (40) are *loose* enough that the feature match satisfies the identity condition and the string is licit, even though such a match would not necessarily be the most natural in a non-elliptical continuation.

Returning to the modal mismatches, let us move on to cases involving the apparent disappearance of modality in the ellipsis site. In these cases, there is a modal in the antecedent and no modal in the sluice:

(42)   *Mismatch Type: apparent disappearance of modality in the sluice* (Rudin 2019)
        Although Sally sees that she **must** defeat her competitors, she relies on Susie to tell her how <**to defeat** her competitors>.

These examples can be handled straightforwardly. The ellipsis site merely lacks the projection hosting the modal, or the ModP projection has no content. Once again, we are dealing with a case where a mismatch is allowed because it instantiates a mismatch between the absence and presence of a feature bundle:

(43)   *Mismatch Type: Disappearance of modality in the sluice*
        Antecedent: $[[_{XP} [_{ModP} \ldots YP]]]$ Mod^0
        Sluice: $[_{XP} \ldots YP]]$ $no\ Mod^0$

A final mismatch type related to modals is observed below. These examples are characterized in Rudin 2019 as involving "abstraction" of modality, a name that highlights the vague nature of the modal that appears to be interpreted in the target clause:[79]

(44)   *Mismatch Type: Abstraction of modality in the sluice* (Rudin 2019)
        Sally said that customers **should** be given lower rates, but Susie said it's hard to see how <customers **could** be given lower rates>.

While this analysis of the example in Rudin 2019 gives rise to a modal mismatch, there is another possible source for the sluice, which I give below:

---

[79] To quote directly: "The case in (23c) [***author note***: the abstraction of modality mismatches], as well as the cases in (23a) and (21b), illustrate a common property of these left-peripheral mismatches under sluicing: their interpretation is somewhat nebulous." (Rudin 2019: 267)

(45)   *Mismatch Type: Abstraction of modality in the sluice (additional interpretation)*
Sally said that customers **should** be given lower rates, but Susie said it's hard to see how <customers **are to be given** lower rates>.

Were we to take the interpretation in Rudin 2019 of the modal content of the sluice at face value (44), we would be forced to conclude that these examples involve a clash of ModPs between antecedent and sluice, thus constituting a problem for the approach to identity defended here. However, (45) could be the underlying source for the sluice, in compliance with featural non-distinctness.

To summarize this subsection, I assessed a range of examples from Rudin 2019 involving mismatches in modality—(i) appearance of modality in the ellipsis site, (ii) disappearance of modality in the ellipsis site, and (iii) abstraction of modality in the ellipsis site. I evaluated the data and concluded that they are all correctly ruled in by the identity condition proposed in this dissertation. In certain cases, we observed that alternative contents to those posited in Rudin 2019 are plausible for the ellipsis sites. These alternative interpretations of the examples comply with the condition proposed in the present work, explaining the well-formedness of the examples. I will return to other examples from Anand et al. 2021 that seem to show modal mismatches in 3.6, at the conclusion of this chapter.

3.3.3 Illocutionary mismatches

Another group of examples discussed in Rudin 2019 involves illocutionary mismatches. For instance, observe that in the example below, the antecedent is an imperative, whereas the target clause is a declarative:

(46)   *Mismatch Type: Illocutionary mismatch* (Rudin 2019)
Always **save** a little from each paycheck. Once you're older, you'll understand why <**you should** always save a little from each paycheck>.

At least some English speakers judge that a sentence without a modal is possible here as the underlying source for the sluice:

(47) *Mismatch Type: Illocutionary mismatch (additional interpretation)*
      Always **save** a little from each paycheck. Once you're older, you'll understand why <**you** always **save** a little from each paycheck>.

Therefore, the assumption that there is a modal *should* in the ellipsis site in (46) cannot be used as the premise in an argument against the identity condition proposed in the present work.

We can focus exclusively on the illocutionary mismatch, then. Analyzing this type of example will depend crucially on our particular approach to the featural content of different clause types (see van der Wurff 2007 for a discussion of different approaches to imperatives). A recent proposal takes imperatives to contain a JussiveP projection immediately above TP that is absent in declaratives (Zanuttini et.al. 2012; see also Isac 2015 for a proposal that is similar in spirit). A piece of evidence for this proposal is that second person pronouns can be bound in imperatives (e.g., 'Wash yourself!'). Zanuttini et.al. propose that the second person feature that binds the pronoun is introduced in the Jussive head. Were one to assume that illocutionary force is encoded in a head within the A'-domain (see the discussion below in 3.4), the fact that second person pronouns can be bound in imperatives would be unexpected, since elements in the A'-domain are not possible antecedents for binding.

We will not delve into the details of this analysis, but highlight instead the main takeaway in relation to the discussion at hand: if imperatives are syntactically encoded via a projection that is absent in declaratives, then Rudin's illocutionary mismatches instantiate (once again) a case where there is a mismatch between the presence and absence of a head

186

bearing a feature bundle. Since this mismatch type is allowed by the approach to identity in ellipsis proposed here, these examples are unproblematic:[80]

(48)   *Mismatch Type: Illocutionary mismatch*
        Antecedent: [$_{JussiveP}$ [$_{XP}$ … YP]]   *Jussive^0; imperative*
        Sluice: [$_{XP}$ … YP]   *no Jussive^0; declarative*

To summarize, the illocutionary mismatches discussed in Rudin 2019 can be seen to comply with the identity condition predicated on featural non-distinctness proposed in this dissertation and are correctly ruled in.

3.3.4 Polarity mismatches

A final type of mismatch that seems to be allowed in ellipsis involves polarity features. These examples are particularly important in showing that the identity condition proposed in this dissertation is on the right track. To begin, consider the example below. The antecedent clause is affirmative, whereas the sluice contains negation:

---

[80] Murphy (2016), in fact, suggests an analysis of this nature to account for this type of mismatch. Murphy proposes a condition on ellipsis where the features of the ellipsis site must be a proper subset of the features of the antecedent:

(i)   *Subset Condition on Ellipsis* (Murphy 2016)
       The morphosyntactic features in the ellipsis site must be a proper subset of those in the antecedent ($F_E \subset F_A$).

Observe, however, that (i) undergenerates: examples involving polarity mismatches (section 3.3.4) and the appearance of modality in the ellipsis site (3.3.2) argue against the condition, and so does our analysis of AF-Active mismatches in Kaqchikel (chapter 2) and certain [GENDER] mismatches (Class II and III masculine-feminine; chapter 4); see chapters 4 and 5 for more comments. In brief, a subset condition on ellipsis is still too strict, since we observe that there can be features in the ellipsis site that are not present in the relevant portion of the antecedent clause.

(49)   *Mismatch Type: Polarity mismatch* (Rudin 2019)[81]

Either **turn in** your final paper by midnight or explain why <you **didn't** turn it in by midnight>!

An additional example showing the same kind of mismatch is shown below, from Kroll 2019:

(50)   *Mismatch Type: Polarity mismatch (Kroll 2019)*

I don't think that California **will comply**, but I don't know why <California **won't comply**>.

In order to account for data like (49)-(50), let us adopt the view that clauses containing negation project a ΣP phrase where the Σ^0 hosts a [+NEG] feature. Conversely, ΣP is absent in affirmative clauses (Laka 1990, 1991). Adopting this analysis, this type of data falls straightforwardly under the proposal defended here: data like (49) involve a mismatch between the absence and presence of a head bearing a feature bundle, in this case, $\Sigma_{[+NEG]}$:

(51)   *Mismatch Type: Polarity mismatch*

| Antecedent: | $[_{XP} \dots YP]$ | *no* Σ^0 |
| Sluice: | $[_{\Sigma P} [_{XP} \dots YP]]$ | $\Sigma^0_{[+NEG]}$ |

Similarly to our discussion of modality mismatches, even if canonical affirmatives did project a polarity projection, it would be featurally empty and thus would not create a featural clash (on the issue of *verum* focus, see (55) and surrounding discussion below).

One might worry at this juncture that the identity condition proposed in this dissertation leads to overgeneration when it comes to polarity mismatches: certain examples seem to comply with featural non-distinctness but are nevertheless ill-formed (Gesoel Mendes p.c.). Consider the following:

(52)   *I know who came to the party, but I don't know who.

---

[81] Example (49)a is attributed to Kroll 2018 in Rudin 2019, but I was not able to track down the example.

First of all, note that this example is judged as unacceptable since it seems to be interpreted as an outright contradiction. What this means is that speakers must be computing the sentence as involving a polarity *match*:

(53) *I know who came to the party, but I don't know who <**came to the party**>.

The question that arises is why we cannot interpret the sluice as involving negation. In other words, the ellipsis site would contain a $\Sigma^0_{[+NEG]}$ that is not present in the antecedent, satisfying featural non-distinctness:

(54) *I know who came to the party, but I don't know who <**didn't** come to the party>.
unavailable

The issue with (54) seems to be that the negation in the ellipsis site carries contrastive focus. As we noted in chapter 2, it has been argued that focused elements cannot be elided (Merchant 2001, Weir 2014, Bennett et al. 2019). There are therefore independent reasons why examples like (52) with the intended interpretation in (54) are ill-formed, even if featural non-distinctness is satisfied. Note, furthermore, that in well-formed examples like (50), negation in the ellipsis site appears to be backgrounded given the content of the preceding overt material.

The polarity mismatches we have just discussed are crucial as an empirical testing ground that can adjudicate between the proposal in this dissertation and alternatives like Rudin 2019. Recall that our discussion of Spanish tense clashes in 3.2 and the Kaqchikel voice mismatches in chapter 2 led us to conclude that the eventive core is *not special in ellipsis*: a partitioning of the clause into material that enters into the calculation of identity (the eventive core) versus material that is exempt from the calculation of identity (elements above the eventive core) cannot be correct. Thus, the availability of polarity mismatches is particularly troubling for an identity condition predicated on strict head-to-head matching

like Rudin 2019 or Merchant 2013a (see chapter 2). In a nutshell, the empirical landscape forced us to conclude that the eventive core is not privileged in relation to the identity condition on ellipsis. A strict identity condition, then, would have to be enforced throughout the clausal spine—there is no hope for the proposal that the clause is partitioned into a section that enters into the condition and another that does not. Why, then, are polarity mismatches allowed? Non-distinctness seems to provide the correct explanation: as long as $\Sigma^0_{[+\text{NEG}]}$ does not clash with a Σ^0 bearing a different polarity feature, then the identity condition is satisfied (modulo the controlling of other factors; see above on negation that is focused in the ellipsis site).

In this context, we can take our discussion beyond the specific examples discussed in Rudin 2019. The approach to the identity condition in this dissertation makes a testable prediction in relation to polarity mismatches. Let us follow Laka 1991 in assuming that emphatic 'do' in English involves a [-NEG] feature on Σ^0. We predict, then, that a clash between the $\Sigma^0_{[\text{-NEG}]}$ and $\Sigma^0_{[+\text{NEG}]}$ under ellipsis should be ruled out as a violation of non-distinctness. As Stockwell & Wong 2020 discuss, this prediction appears to be borne out:

(55)   *Mismatch Type: Polarity clash* (Stockwell & Wong 2020)
   a. Either he turned in his final paper by midnight or he explained why <he **didn't** turn it in by midnight>. $\emptyset - \Sigma^0_{[+\text{NEG}]}$
   b. ?? Either he **DID** turn in his final paper by midnight or he explained why <he **didn't** turn it in by midnight>. $\Sigma^0_{[\text{-NEG}]} - \Sigma^0_{[+\text{NEG}]}$

The condition in (32) correctly rules in the examples where there is no clash between Σ heads bearing polarity features (49) and (55)a, while also correctly ruling out data like (55)b that involve a featural clash. In contrast, the proposal in Rudin 2019 would predict that (55)b should be well-formed, since ΣP is outside the eventive core.

To end this subsection, let us consider once more free exceptives (for the distinction between connected and free exceptives, see Hoeksema 1987), which have been analyzed as involving ellipsis in English (Stockwell & Wong 2020; see Harris 1982, Merchant 2001: 107, fn. 12), Spanish (Pérez-Jiménez & Moreno-Quibén 2012; see 3.2 above), Malagasy (Potsdam 2018, 2019; Potsdam & Polinsky 2019), and Egyptian Arabic (Soltan 2016). As Potsdam 2019 emphasizes, there exists a generalization regarding the polarity of main versus exceptive clauses:

(56)   *Polarity Generalization* (from Potsdam 2019; after García Álvarez 2008)
        The propositions expressed in the main clause and exceptive clause must have opposite polarity.

Consider the Malagasy and Egyptian Arabic examples below to illustrate the generalization (we will return to the details of Malagasy ellipsis and sluicing in chapter 5, where we discuss apparent voice mismatches in the language). Observe that the ellipsis site in (57)b contains a negative marker that is not present in the antecedent:

(57)   *Malagasy exceptives* (adapted from Potsdam 2019 ex. (5) & (6)a)
        a. Tonga   ny vahiny rehetra afa-tsy Rasoa no   tsy tonga.
           arrived DET guests all   except Rasoa FOC NEG arrive
           'All the guests arrived, except Rasoa (didn't arrive).'
        b. Tonga   ny   vahiny rehetra afa-tsy Rasoa <no   tsy tonga>.
            arrived DET guests all   except Rasoa   FOC NEG arrive
           'All the guests arrived, except Rasoa.'

The same can be observed in the Egyptian Arabic examples below:[82]

(58)   *Egyptian Arabic exceptives* (adapted from Soltan 2016, ex. (3)a & (28))
        a. ʔanaa šuf-t   kull ʔil-talaba   ʔillaa Ahmad maa-šuf-t-uu-š.
           I   saw-1S all   the-students except Ahmad NEG-saw-1S-him-NEG
           'I saw all the students, except Ahmad I didn't see him.'

---

[82] A perfect minimal pair is not provided by Soltan (2016); note that the elliptical (58)b contains a prepositional phrase 'at the lecture' in the antecedent that is not present in the non-elliptical (58)a.

b. ʔanaa šuf-t   kull ʔil-talaba   fii ʔil-muħaadra ʔillaa   Ahmad.
I   saw-1s all   the-students at the-lecture   except Ahmad
'I saw all the students at the lecture, except Ahmad.'

At first glance, the takeaway from the data in (57) and (58) seems to be clear: if (i) free exceptives like the above involve ellipsis and (ii) there is negation inside the ellipsis site, then these examples constitute an additional case in which a head bearing a negative feature in the ellipsis site mismatches with the absence of a head bearing a polarity feature in the antecedent (see Vostrikova 2019 for arguments that the ellipsis site in exceptives must include negation).[83] In a manner parallel to the polarity mismatches in sluicing (49)-(55),

---

[83] Vostrikova's main argument is as follows. First, except clauses in general force a polarity reversal:

(i)   a. Every girl was there except Eva <was not there>.
        b. No girl was there except Eva <was there>.   (Vostrikova 2019: 422)

With this in mind, Vostrikova uses NPI licensing as a diagnostic for the presence or absence of negation within the ellipsis site. According to Vostrikova, the contrast we observe below is expected if negation is present in the ellipsis site in (ii)a, but not in (ii)b, as shown in (iii)a,b:

(ii)   a. John danced with everyone except with any girls from his class.
        b. *John danced with no one except with any girls from his class.   (Vostrikova 2019: 423)

(iii)   a. John danced with everyone except <John did not dance> with any girls from his class.
        b. *John danced with no one except <John danced> with any girls from his class.
                                                         (adapted from Vostrikova 2019: 423)

In the examples above, the PP modifier 'with any girls from his class' (which contains the NPI 'any') has presumably moved out of the ellipsis site (see Stockwell & Wong 2020). These results in Vostrikova 2019 would seem to be at odds with the conclusions in Stockwell & Wong 2020 that it is the except-phrase which contributes the negative meaning (see the discussion of the contrast in (59)). Stockwell & Wong, however, counter Vostrikova's argument by showing first that an NPI can associate with an except-phrase when said NPI is higher, and thus not c-commanded by, a putative elided negation:

(iv)   a. Everyone came, except any boys from Mary's class.
        b. Everyone came except any boys from Mary's class <**didn't** come>.
                                                         putative negation in ellipsis-site
        c. *Any boys didn't come.   (Stockwell & Wong 2020; ex. (26))

Second, an NPI cannot in general be moved out of the scope of its licensor:

192

then, this type of mismatch in exceptives is correctly ruled in by the identity condition proposed in this dissertation.

However, some of the aforementioned authors have proposed that the source of the feature that gives rise to the negative interpretation in the exceptive clauses is not, in fact, inside the ellipsis site. Instead, the negative feature is either (i) encoded in the *except*-phrase itself (Potsdam 2018, 2019, Potsdam & Polinsky 2019 for Malagasy; Stockwell & Wong 2020 for English; Soltan 2016 for Egyptian Arabic) or (ii) the negative feature is outside of the ellipsis site in an immediately dominating projection—i.e., PolP is above TP, and TP undergoes ellipsis (Soltan 2016).[84]

(v)   a. Sam does not like silly pictures of any of his friends.
      b. *Which pictures of any of his friends does Sam not like *t* ?   (Stockwell & Wong 2020; ex. (28))

If the PP modifier in (ii)a moved out of the scope of its licensor (the putative negation in the ellipsis site), then its well-formedness is unexpected. What Stockwell and Wong do not address, however, is why Vostrikova's (ii)b is ill-formed—presumably, the except-phrase should be able to license the NPI in that case, contrary to fact. All in all, this discussion should highlight that establishing the usefulness of exceptives as a construction that can adjudicate between different formulations of the identity condition will require more detailed work in the future.

[84] A few questions arise here regarding these proposals. In Malagasy, the except-phrase cannot be analyzed synchronically as containing a negative marker (Maria Polinsky p.c.) and the ellipsis site in the exceptive examples clearly contain negation. In Egyptian Arabic, it is unclear why the negative marker does not surface in the free exceptive; Soltan (2016: fn. 12) notes that it is possible that negative Pol^0 (which under his analysis is outside the ellipsis site) incorporates into the except-phrase. However, as acknowledged by the author, there is only diachronic, not synchronic evidence that the exceptive-phrase contains the negative marker *-laa*.

Stockwell & Wong (2020) provide the following examples to support their conclusion that it is the *except*-phrase which contributes the negative meaning in English exceptives. The first pair in (59)a,b below shows the asymmetry we discussed before in the availability of polarity mismatches in sluicing: Rudin 2019 shows that this type of example is well-formed when the antecedent is a simple affirmative (59)a, whereas Stockwell & Wong (2020) show that the example is ill-formed when the antecedent is an emphatic affirmative (59)b. In contrast, the examples in (59)c-d show that a polarity mismatch is well-formed in exceptives even when the antecedent is an emphatic affirmative (data from Stockwell & Wong 2020):

(59)   *A contrast in polarity mismatches; sluicing vs. exceptives*
    a. Either he turned in his final paper by midnight or he explained why <he **didn't** turn it in by midnight>.
    b. ?? Either he **DID** turn in his final paper by midnight or he explained why <he **didn't** turn it in by midnight>.
    c. Everyone liked the movie, except John.
    d. Everyone **DID** like the movie, except John.

Stockwell & Wong (2020) adopt the identity condition defended in this dissertation thus far and interpret the data above as evidence that the negative interpretation of the target clause in the exceptive cases is contributed by the *except*-phrase itself, which is outside of the ellipsis site. There is therefore no head bearing a [+NEG] feature within the ellipsis site that would clash with the head bearing [-NEG] (emphatic affirmative) in the antecedent. If Stockwell & Wong are correct, then English exceptives *by themselves* do not provide evidence in support of the identity condition defended here. However, the contrast in mismatch availability between the sluicing cases (59)a,b and the exceptives (59)c,d *does* provide evidence in support of the identity condition defended in this dissertation.

All in all, exceptives in several languages provide evidence that polarity mismatches are possible under ellipsis. However, the question of the locus of the negative interpretation of the exceptive clause needs to be answered on a language-by-language basis. If the negative interpretation of the target clause in exceptives results from the existence of a head bearing a negative feature inside the ellipsis site, then there is a polarity mismatch in exceptives that complies with featural non-distinctness; this alone would provide evidence for the identity condition proposed in this dissertation. Alternatively, if the negative feature in the target clause of exceptives is *not* inside the ellipsis site, then exceptives provide support for the identity condition only insofar as language internal contrasts like the English (59)a-d arise.

To summarize this subsection, I first discussed the sluicing examples described in Rudin 2019 as involving a polarity mismatch and concluded that they are correctly ruled in by the identity condition proposed in this dissertation. I also discussed an asymmetry brought to light by Stockwell & Wong 2020 which provides support for the proposal here: whereas a simple affirmative in the antecedent can mismatch with negation in the ellipsis site, an emphatic affirmative (*verum* focus) in the antecedent cannot. I followed Stockwell & Wong 2020 in analyzing the latter (ill-formed) examples as instantiating a clash between Σ heads bearing [-NEG] and [+NEG] features, a configuration that violates featural non-distinctness. After discussing the sluicing examples, we moved on to exceptives and noted how this elliptical construction seems to provide additional support for the identity condition proposed here. However, some questions remain about the locus of the negative feature in exceptives cross-linguistically. The results in this subsection, in combination with our conclusion in 3.2 that the eventive core is not special in the computation of identity

195

in ellipsis, show that syntactic identity conditions like those proposed in Rudin 2019 and Merchant 2013a are untenable.

3.3.5 Mood mismatches

A final type of example that is not analyzed in Rudin 2019 involves mood mismatches in Spanish stripping. Consider the data below, where the antecedent is in the imperative mood, while the ellipsis site is subjunctive:

(60)   *Spanish stripping—mood mismatch* (see Saab 2003, 2008, 2021)
    a. Ahorrá   plata,   no palabras$_1$ <[$_{TP}$ ahorrés t$_1$]>.
       save.IMP money not words   save.SUBJ
       'Save money, not words!'   (From an Argentine commercial)
    b. No ahorrés   plata,   pero sí   palabras$_1$ <[$_{TP}$ ahorrá t$_1$]>.
       not save.SUBJ money but   yes words   save.IMP
       'Don't save money, save words!'   (approximate translation)

As laid out in Saab 2021, a way of analyzing these examples (and which is fully consistent with the identity condition proposed here) is that the mood mismatch is just an illusion and there is no syntactic mismatch here between the antecedent and ellipsis site.[85] Saab discusses Harris (1998)'s analysis, which posits that the syntax of imperatives and subjunctives in Spanish is identical; the difference in spell-out of the relevant verbs that display "imperative" or "subjunctive" morphology is the result of a morphological impoverishment rule that is post-syntactic. The featural content of imperatives and subjunctives under that proposal is therefore identical, and thus examples like (60) satisfy featural non-distinctness.

---

[85] Icelandic case mismatches seem to be another empirical domain where an analysis of this sort seems promising; see Wood et al. 2019.

An additional type of example that can be handled in a similar manner involves mood mismatches of the following kind in Spanish: here, the antecedent verb is in the indicative mood, whereas the verb in the ellipsis site shows subjunctive:

(61)　　*Mood mismatch in Spanish—indicative – subjunctive* (adapted from Brucart 1999)
　　　　Juan no irá　　a la　fiesta, pero es posible　que María sí　<vaya
　　　　Juan not go.FUT to the party　but　is　possible that María yes　go.SUBJ
　　　　a la　fiesta>.
　　　　to the party
　　　　'Juan will not go to the party, but it's possible that María will.'

This example can be handled straightforwardly if Mood features are generated on C^0 (hence outside the ellipsis site) and lower onto T^0. One way of analyzing this would be via Feature Inheritance (Chomsky 2008; see Gallego 2014).[86]

3.3.6 Taking stock of mismatches above Voice

Let us summarize our conclusions from section 3.3. I evaluated data discussed in Rudin 2019 involving different types of mismatches that are allowed in English sluicing: tense mismatches (3.3.1), modality mismatches (3.3.2), illocutionary mismatches (3.3.3), and polarity mismatches and (3.3.4). Along the way, additional data from Spanish, Malagasy, and Egyptian Arabic were brought in to supplement the discussion. I showed that all of the data can be accounted for via the identity condition defended in this dissertation. In subsection 3.3.5, I briefly discussed mood mismatches in Spanish, concluding that these do not pose an issue to the identity condition. The table below summarizes the data

---

[86] The data discussed in this subsection, as well as the explanation for the well-formedness of (60)-(61), were suggested by an anonymous reviewer for earlier versions of the material discussed in this chapter (Ranero 2019b). I leave for the future an assessment of whether the analysis of Spanish imperatives suggested here could be applied to the English illocutionary mismatches discussed earlier.

discussed in this chapter so far, showing how mismatches that satisfy featural non-distinctness are correctly ruled-in by non-distinctness under ellipsis, whereas those that involve featural clashes (and thus violate featural non-distinctness) are correctly ruled out:[87]

(62)   *The data so far (mismatches above Voice)*

Status	Antecedent	Ellipsis site	Language	Example
*	$T_{PRESENT}$	T_{PAST}	Spanish	(17)B', (27)b
*	T_{PAST}	$T_{PRESENT}$	Spanish	(18)B'
*	T_{PAST}	T_{FUTURE}	Spanish	(21)d
*	T_{FUTURE}	T_{PAST}	Spanish	(33)d ch. 1
*	historical $T_{PRESENT}$	T_{PAST}	Spanish	(26)
*	T_{PAST}	$T_{PRESENT}$	English	(34)B′
*	$\Sigma_{[-NEG]}$	$\Sigma_{[+NEG]}$	English	(55)b
✓	$T_{PRESENT}$	Tø (infinitival)	English	(29)a
✓	Tø (infinitival)	$T_{PRESENT}$	English	(29)b, (38)
✓	$T_{PRESENT}$	Tø + modal *will*	English	(29)c
✓	Tø (gerund)	Tø (infinitival)	English	(30)
✓	Ø	Mod^0	English	(35)
✓	Mod^0	Ø	English	(45)
✓	$Juss^0$	Ø	English	(47)
✓	Ø	$\Sigma_{[+NEG]}$	English	(49), (50), (55)a
✓	Ø	$\Sigma_{[+NEG]}$	Spanish	(27)a
✓	Ø	$\Sigma_{[+NEG]}$	Malagasy	(57)
✓	Ø	$\Sigma_{[+NEG]}$	Egyptian Arabic	(58)

[87] Note that we should only interpret the exceptive examples in Spanish, Malagasy, and Egyptian Arabic as well-formed mismatches if negation is contributed by an element *inside* the ellipsis site (Vostrikova 2019), rather than being contributed by the except-phrase which is outside of the ellipsis site (Stockwell & Wong 2020; Potsdam 2018, 2019, Polinsky & Potsdam 2019; Soltan 2016; see 3.3.4 for discussion on these competing analyses). If for an example of this type, the latter analysis proves to be correct, then that example is uninformative for our purposes and cannot adjudicate between the identity condition proposed in this dissertation and others like Rudin 2019's.

Readers might wonder at this juncture whether all the relevant mismatches discussed for English can be constructed in Spanish as well, and whether they behave like English. As far as I can tell, some of the modality mismatches result in the same variation in interpretive possibilities and there is no diagnostic to ensure which, if any, modal expression must occur inside the sluice. The illocutionary mismatches do not result in ill-formedness, much like in English. Below we observe an imperative antecedent and a declarative or a bare infinitival in the sluice; the example is a translation of the English case we assessed in (46)-(47):

(63) *Illocutionary mismatch in Spanish (sluicing)*
      Siempre **guardá**_{IMP} algo de cada cheque! Cuando seas grande, entenderás por qué
      <**guardás**_{DECL} / ?**guardar**_{INF} algo de cada cheque>.

The possibility of polarity mismatches parallel to the sluices discussed in Rudin 2019, Kroll 2019, and Stockwell & Wong 2020 is less clear at present. At least according to my judgments and those of my consultants, the negative marker is preferable outside of the ellipsis site. The examples below are a direct translation of the English data in (49):

(64) *Polarity mismatch in Spanish (sluicing)*
      a. ??O entregó   su ensayo final a medianoche, o   explicó   por qué <no
         or turned.in.3s her essay   final at midnight   or explained.3s why   not
         entregó   su ensayo final a medianoche>.
         turned.in.3s her essay   final at midnight
      b. O entregó su ensayo final a medianoche, o explicó por qué no <entregó su ensayo
         final a medianoche>.

Note that example (64)a complies with feature non-distinctness, but is nevertheless deviant. I leave for the future a more thorough exploration of why an asymmetry exists between Spanish and English examples that appear on the surface to be equivalent—perhaps the relative height of negation with respect to the ellipsis site can provide a way to account for this apparent difference.

I will proceed with the task of expanding the empirical domain further still in section 3.5. However, before doing so, the following section briefly discusses a potential alternative analysis for some of the data discussed above and in Rudin 2019.

3.4 An alternative analysis for some mismatches

A question that arises now is whether an alternative explanation could be offered for the well-formedness of some of the examples that we have analyzed as involving mismatches that comply with featural non-distinctness. For example, one could pursue an analysis proposing that, in a subset of the data we have assessed so far, the relevant features that mismatch are actually outside of the ellipsis site (e.g., see the above discussion of exceptives and polarity mismatches in 3.3.4). A finiteness mismatch in Spanish sprouting will allow us to illustrate this approach; these data are similar to the English example in (29)a.:

(65)   *Spanish sprouting—finiteness mismatch allowed* (Saab 2021)
  a. Recuerdo haber arreglado el carro, pero no recuerdo cuándo
   remember.1SG have.INF fixed the car but not remember.1SG when
  <arreglé el carro>.
   fixed.1SG the car
  'I remember having fixed the car, but I don't remember when.'

  b. Juana finalmente arregló el carro, aunque parecía no saber cómo
   Juana finally   fixed.3S the car although seemed not know.INF how
   <arreglar el carro>.
    fix.INF the car
  'Juana finally fixed the car, although she didn't seem to know how.'

If one were to assume (i) an expanded left-periphery (Rizzi 1997 and many others) and (ii) that finiteness is encoded on FinP (outside of the ellipsis site), then these mismatches are handled straightforwardly (see Saab 2021, Tanaka 2011 for explanations of this kind). If this analysis is on the right track, then examples like (65), would not support the proposal

in Rudin 2019 that material outside the eventive core does not enter into the calculation of identity in ellipsis. In fact, an analysis of this sort would lead us to conclude that finiteness mismatches cannot serve as empirical testing ground to adjudicate between a syntactic identity condition like Rudin 2019's and the proposal in this dissertation.

A similar tactic could be used for the illocutionary mismatches. In other words, if one were unconvinced by Zanuttini et al. 2012's argumentation for a distinction between imperatives and declaratives based on the presence/absence of a JussiveP projection, one could assume instead that force is encoded higher in the left periphery (e.g., in Rizzi 1997's ForceP; see, however, 3.3.3 for reasons to question the viability of this approach). Under this analysis, once again, one would posit that the relevant mismatching features are outside of the ellipsis site, so they would not enter into the calculation of identity.

It is not unreasonable to expect that some cases could be explained by proposing a fine-grained distribution of features in the left-periphery and arguing that the size of the ellipsis site is not what it seems at first glance. In light of this discussion, however, it is important to make clear that this line of analysis—attempting to locate the relevant mismatching feature outside of the ellipsis site for every case in which there exists a well-formed mismatch—is much more difficult to maintain in some of the cases that we have discussed. For instance, I fail to see how this line of analysis could account for mismatches where it appears inevitable to assume that the mismatching features are inside of the ellipsis site (e.g., modality (3.3.2) and polarity (3.3.4) mismatches under sluicing). Furthermore, this approach cannot derive the subset of voice mismatches that are allowed under sluicing in Kaqchikel (chapter 2). Finally, as we will see in chapter 4, this approach cannot capture

the totality of the data related to well-formed and ill-formed [GENDER] mismatches in the nominal domain.

In the next subsection, I extend the empirical coverage of the identity condition proposed here further, looking at [PERSON] mismatches that, initially, appear to be problematic for my approach.

3.5 Extending the empirical domain: [PERSON] mismatches and proper containment

So far in this chapter, we have shown that there exists converging evidence for an identity condition on ellipsis that is calculated on the basis of featural *non-distinctness*, as opposed to featural *identity*. In short, the contrast between tense mismatches in English and Spanish, coupled with our Kaqchikel results in chapter 2 and our re-evaluation of the mismatches discussed in Rudin 2019, all provide support for the identity condition proposed in this dissertation.

However, there is an additional type of mismatch that might, at first glance, appear to be problematic for the proposal in the present work. In the examples below involving TP-ellipsis in Spanish, the verb in the antecedent inflects for a different person specification than the verb inside the ellipsis: in short, there appears to be a clash between φ-features. However, the examples are licit, contra expectations:[88]

---

[88] Saab 2008: 264-267 provides a different interpretation of the well-formedness of the inflectional mismatches discussed here. Saab argues that these examples can be handled if (i) strict identity is evaluated at LF and (ii) inflectional features like [PERSON] are deleted by the Agree operation. Therefore, the featural representation that is evaluated by the identity condition is one where there is no [PERSON] mismatch, since these features are not present at LF.

(66)   *TP-ellipsis in Spanish—person inflection mismatch*

    a. José **fue** al parque, pero Julián no <**fue** al parque>.   ✓3-3
      José **go.3S.PST** to.the park   but Julián NEG **go.3S.PST** to.the park
      'José went to the park, but Julián did not.'

    b. José **fue** al parque, pero yo no <**fui** al parque>.   ✓3-1
      José **go.3S.PST** to.the park   but 1s NEG **go.1S.PST** to.the park
      'José went to the park, but I did not.'

    c. José **fue** al parque, pero vos no <**fuiste** al parque>.   ✓3-2
      José **go.3S.PST** to.the park   but 2s NEG **go.2S.PST** to.the park
      'José went to the park, but you did not.'

In (66)b-c, 3rd person in the antecedent mismatches with 1st or 2nd person in the target clause: the T^0 head in the antecedent has entered into an Agree relation with the subject of the antecedent, bearing as a result the subject's φ-features; similarly, the T^0 head in the target clause has entered into an Agree relation with the subject of that clause—the T^0 head in the antecedent and the T^0 head in the ellipsis site thus mismatch.

To begin assessing our options, there is one simple way that the examples above could fit into our proposal as it currently stands. If we considered 3rd person to actually be the *absence* of person features (Benveniste 1971, Harley & Ritter 2002, McGinnis 2005), then these mismatches would be of the same nature as others we have previously discussed, since one of the T^0 heads would bear no person features at all.

This analysis cannot cover the full range of person mismatches that are well-formed, however, as evidenced by the examples below. In (67)a,b, there is a mismatch between 3PL in the antecedent 1S or 2S in the ellipsis site. If we follow the proposals in Harley & Ritter 2002 and McGinnis 2005, 3PL and 1S person do not stand in a subset relation, and 3PL and 2S do not stand in a subset relation either. In other words, we have a case here where T^0 in the antecedent bears the φ-feature bundle of the 3PL pronoun, whereas T^0 in the ellipsis site bears the feature bundle of a local person—1S in (67)a and 2S in (67)b.

(67) *TP-ellipsis in Spanish—person feature mismatch*

    a. Ellos **fueron** al parque, pero yo no <**fui** al parque>.
       3P   **go.3P.PST** to.the park   but 1S NEG **go.1S.PST** to.the park
       They went to the park, but I did not.'   ✓3P-1S

    b. Ellos **fueron** al parque, pero vos no <**fuiste** al parque>.
       3P   **go.3P.PST** to.the park   but 2S NEG **go.2S.PST** to.the park
       'They went to the park, but you did not.'   ✓3P-2S

Abstractly, the kind of configuration we are faced with here is different from others that have been assessed thus far in this dissertation. This seems to be a case where a head H^0 in the antecedent and the corresponding H^0 in the ellipsis site do not bear the same feature bundle, but these features are not directly related—in other words, H^0 in the antecedent bears a feature [F], while H^0 in the ellipsis site bears a feature [G]:

(68) *Mismatch in the abstract*
    Antecedent: $H_{[F]}$
    Ellipsis site: $H_{[G]}$

I will assume that this kind of configuration is a violation of the identity condition—the two heads H^0 here give rise to a featural clash. To make explicit the problem, let us follow Harley & Ritter's 2002 representation of person features and assume that a mismatch between 3PL and 1S constitutes a configuration where T^0 in the antecedent bears [GROUP] and T^0 in the ellipsis site bears [PARTICIPANT], while a mismatch between 3PL and 2S constitutes a configuration where T^0 in the antecedent bears [GROUP] and T^0 in the ellipsis site bears [PARTICIPANT, ADDRESSEE] (see McGinnis 2005 for an elaboration of Harley & Ritter's proposal):

(69) *Person mismatches*
    a. *Antecedent*: $T_{[GROUP]}$ *Ellipsis site*: $T_{[PARTICIPANT]}$ ✓3PL-1S (67)a
    b. *Antecedent*: $T_{[GROUP]}$ *Ellipsis site*: $T_{[PARTICIPANT, ADDRESSEE]}$ ✓3PL-2S (67)b

Even though we would expect that the featural clashes sketched above would violate the identity condition and give rise to ill-formed ellipses, these examples are ruled in.

The solution that I propose for this puzzle capitalizes on one key difference between the person mismatches just discussed, which are well-formed, and ill-formed examples like the tense clashes in Spanish (see 3.2 above). First, recall the identity condition defended thus far and focus on the proper containment portion of the proposal:

(70) *Syntactic identity in ellipsis (repeated from (1))*
      Antecedent and material **properly contained within the ellipsis site** must be featurally non-distinct.

Let us define proper containment as follows:

(71) *Proper containment in ellipsis*
      X is properly contained in Y if there is at least one occurrence of Y that dominates every occurrence of X.

What the proper containment portion of the condition does is exempt an element that is not properly contained within the ellipsis site from giving rise to violations of the identity condition (see below).

In a nutshell, I propose that there exists a crucial asymmetry between the well-formed and ill-formed examples regarding proper containment. For the tense clashes, the clashing features are properly contained within the ellipsis site, as shown by the schematic below:[89]

---

[89] The details of where the subject sits in the antecedent in (72)b are irrelevant for our purposes, but I place it in SpecTP for ease of exposition; see Ordóñez (1997).

(72)   *Proper containment of clashing features; Tense mismatch* (repeated from (17)b)

a. *Hoy, Ana **repara** los celulares, pero ayer, no sé quién <**reparó** los celulares>.
b. *Antecedent*: [$_{CP}$ [C [$_{TP}$ [**T**…]]]]
c. *Ellipsis*:   [$_{CP}$ [C$_{[E]}$ [$_{\text{TP}}$ [**T**…]]]]

In contrast, this is not the state of affairs for the person mismatches: one of the bearers of

the relevant features (the pronoun controlling agreement) is a remnant outside of the ellipsis

site:[90]

(73)   *No proper containment; person mismatch (repeated from (67)a)*

a. Ellos **fueron** al parque, pero **yo** no <**fui** al parque>.
b. *Antecedent*: [$_{CP}$ [C [$_{TP}$ **DP$_{SUBJ}$** [**T+V**…]]]]
c. *Ellipsis*:   [$_{XP}$ **DP$_{SUBJECTi}$** X [$_{ΣP}$ Σ$_{[E]}$ [$_{\text{TP}}$ **DP$_i$** [**T+V**…]]]]

Why should this asymmetry be relevant? To answer this, consider a separate empirical

domain—the merger and sprouting types of sluicing. The *merger* type of sluicing involves

sluicing where the *wh*-remnant in the target clause has a correlate in the antecedent;

---

[90] The proposal that the notion of proper containment within the ellipsis site is relevant to the identity

condition on ellipsis is found in Preminger 2011. Rudin 2019's proposal encodes a similar intuition. Consider

the following example, where the object in the ellipsis site is ignored:

(i)   Sally ate, but I don't know **what** <Sally ate>. (Rudin 2019)

Commenting on the example above, Rudin 2019: 260 states: "*The antecedent and the ellipsis site are not*

syntactically isomorphic to each other; the difference lies in the presence of a structural object position in

the ellipsis site. The content of that object position, however, is a trace (or unpronounced lower copy), which

*is to say **the tail of a movement dependency chain** [author note*: emphasis mine]; *therefore, the rule in (8)*

[author's note: a preliminary version of the ultimate identity condition proposed in the paper] *does not*

require it to have a structure-matching correlate."

sprouting, meanwhile, involves sluicing where the *wh*-remnant in the target clause does not have a correlate in the antecedent (see Chung et al. 1995):[91]

(74)   *Merger and sprouting types of sluicing*

    a. Sebas went skiing **somewhere**, but he doesn't remember **where** \<he went skiing **where**$_1$\>.   *merger-type of sluicing*

    b. Sebas went skiing, but he doesn't remember **where**$_1$ \<he went skiing **where**$_1$\>.   *sprouting-type of sluicing*

Now, any theory of identity under ellipsis must face the following problem raised by sluicing and any instance of ellipsis where a movement dependency is launched from the ellipsis site: why is it that in calculating identity, the copy of a *wh*-remnant inside of the ellipsis site is ignored? I propose that the proper containment clause in the identity condition does the necessary work: elements that are not properly contained within the ellipsis site are exempt from the calculation of identity in ellipsis. In sluicing, notice that the *wh*-remnant in the target clause is the highest link of a *wh*-movement chain which is not properly contained within the ellipsis site. Therefore, the presence of one (or more) copies within the ellipsis site is ignored.[92]

---

[91] To be more precise about this terminology, "merger" and "sprouting" referred to distinct processes in the LF-copying approach to sluicing defended by Chung et al. (1995). These names, however, have now become synonymous with the constructions themselves (see Chung et al. 2011).

[92] A question arises here in relation to verb-stranding VP ellipsis in languages like Russian, where the verb remnant (which is part of a head-movement chain) in a target clause must be identical in some cases with a verb in the antecedent (Gribanova 2018; cf. Gribanova 2013). The fact that strict identity is enforced between the verbs in this ellipsis type might challenge the proposal advanced here that members of movement chains launched from an ellipsis site are ignored by the identity condition if they have a landing site outside of the ellipsis site. I cannot delve into the intricacies of the verb-stranding VP ellipsis facts here but refer the reader to Saab 2021 and references therein for a discussion of the facts. A proposal in Schoorlemmer & Temmerman

Let us now return to [PERSON] features and see how the proper containment portion of the identity condition accounts for the well-formedness of the mismatches in (66)-(67). I adopt the proposal that φ-feature agreement resulting from the operation Agree is the result of feature sharing (Pesetsky & Torrego 2001, 2007, Frampton & Gutmann 2000, 2006, Chung 2013b). What this means is that in cases of φ-feature agreement between a DP and T^0, DP and T^0 *share* a set of φ-features, rather than DP *assigning* features to T^0. In other words, the post-agreement representation is one where what is literally a single object— the φ-feature bundle—is linked to two other syntactic objects: the DP and the φ-probe (T^0):

(75)   *Agreement is feature-sharing*
     [$_{CP}$ [C [$_{TP}$ DP   [T+V …]]]]
                    └── [φ]──┘

In person mismatches, one of the person feature sharers in the target clause is a remnant outside of the ellipsis site. Given this, the person feature bundle is not properly contained within the ellipsis site. As a result of (70), the elements participating in the feature-sharing relation (i.e., T+V within the ellipsis site) are exempt from the calculation of identity. Just like in cases of sluicing (where the lowermost *wh*-copy is exempt for the purposes of (70)), person features are ignored here as well—note that I use 3PL and 1S as shorthand for the relevant feature bundles.

(76)   *Person features are exempt; there is no mismatch*
     Ellos **fueron** al parque, pero **yo** no <**fui** al parque>.   (repeated from (67)a)
     Antecedent   *Ellipsis site*
     [$_{CP}$ [$_{TP}$ **DP**   [**T+V**…]]] [$_{XP}$ **DP$_i$** X [$_{ΣP}$ Σ$_{[E]}$ [$_{TP}$ ~~DP$_i$~~ ~~[T+V~~…]]]
             └─[φ: 3PL]─┘   └────[φ: 1S]────────┘

_____

2012 might provide a solution—if (i) head movement occurs at PF and (ii) the identity condition is calculated at LF, then for the purposes of the calculation of identity, the relevant verb is properly contained within the ellipsis site and thus must be identical to a verb in the antecedent.

We can now understand why person mismatches of this kind are licit under ellipsis: unlike in tense clashes, person features in these configurations are not properly contained within the ellipsis site. As a result, they do not enter into the calculation of syntactic identity at all.

Let us summarize this subsection and the lessons stemming from it. I discussed an additional type of mismatch, observing that it seemed at first sight to create a problem for our proposal: unlike tense features (see 3.2), clashes in person features result in well-formed examples. To account for this, I capitalized on the proper containment portion of the identity condition and adopted the analysis of φ-feature agreement resulting from Agree as feature sharing and thus derived the possibility of person mismatches: since person mismatches in the target clause involve a remnant DP outside of the ellipsis site that participates in a feature-sharing relation with an element inside the ellipsis site, then those features are ignored altogether. In other words, I proposed the following: when only some, and not all, of the feature sharers are properly contained within the ellipsis site, then those features are exempt from the calculation of identity under ellipsis.[93]

3.6 Conclusion, recapitulation, and final comments

This chapter provided converging evidence in support of the identity condition proposed thus far in this dissertation:

(77)   *Syntactic identity in ellipsis (repeated from (1))*
        The antecedent and material properly contained within the ellipsis site must be featurally non-distinct.

---

[93] Cases involving "vehicle change" (Fiengo & May 1994) and indexical mismatches are also interesting in this context; these are discussed in chapter 5, which lays out several issues deserving of future exploration.

The general strategy that was taken in order to provide support for (77) was to compare its predictions to those made by a different syntactic identity condition (Rudin 2019) that cleaves the clause into an area that enters into the calculation of identity—the eventive core (VoiceP/vP)—and an area that does not. The proposal in Rudin 2019 sought to derive an asymmetry between mismatches involving material inside the eventive core—all ill-formed in English—and mismatches involving elements above it—all (seemingly) well-formed in the language. I introduced a range of cross-linguistic data and showed that the approach taken in Rudin 2019 cannot be the correct strategy to explain the mismatch asymmetry. Contrary to the predictions of such an approach, for example, tense clashes in Spanish are impossible under ellipsis. I then reviewed other data from Rudin 2019 and showed that they are all ruled-in correctly by my proposal. The table below summarizes the data discussed in this chapter, alongside the data we analyzed in chapter 2. The red cells exemplify featural clashes, whereas the green exemplify mismatches that satisfy featural non-distinctness and are correctly ruled in:

(78) *The data so far (all mismatches)*

Status	Antecedent	Ellipsis site	Language	Example
*	$\text{Voice}_{\text{PASS}}$	$\text{Voice}_{\text{ACT}}$	English, Spanish...	(3) ch. 2
*	$\text{Voice}_{\text{ACT}}$	$\text{Voice}_{\text{PASS}}$	English, Spanish...	(2)
*	Voice_{AP}	$\text{Voice}_{\text{ACT}}$	Kaqchikel	(75)-(77) ch. 2
*	T_{PRESENT}	T_{PAST}	Spanish	(17)B', (27)b
*	T_{PAST}	T_{PRESENT}	Spanish	(18)B'
*	T_{PAST}	T_{FUTURE}	Spanish	(21)d
*	T_{FUTURE}	T_{PAST}	Spanish	(33)d ch. 1
*	historical T_{PRESENT}	T_{PAST}	Spanish	(26)
*	T_{PAST}	T_{PRESENT}	English	(34)B'
*	$\Sigma_{\text{[-NEG]}}$	$\Sigma_{\text{[+NEG]}}$	English	(55)b
✓	$\text{Voice}_{\text{ACT}}$	Ø (AF)	Kaqchikel	(78) ch. 2
✓	Ø (AF)	$\text{Voice}_{\text{ACT}}$	Kaqchikel	(79)-(81) ch. 2

✓	Voice$_{PASS}$	Ø (AF)	Kaqchikel	(82),(136) ch.2
✓	T$_{PRESENT}$	T$_\emptyset$ (infinitival)	English	(29)a
✓	T$_\emptyset$ (infinitival)	T$_{PRESENT}$	English	(29)b, (38)
✓	T$_{PRESENT}$	T$_\emptyset$ + modal *will*	English	(29)c
✓	T$_\emptyset$ (gerund)	T$_\emptyset$ (infinitival)	English	(30)
✓	Ø	Mod^0	English	(35)
✓	Mod^0	Ø	English	(45)
✓	Juss^0	Ø	English	(47)
✓	Ø	Σ$_{[+NEG]}$	English	(49), (50), (55)a
✓	Ø	Σ$_{[+NEG]}$	Spanish	(27)a
✓	Ø	Σ$_{[+NEG]}$	Malagasy	(57)
✓	Ø	Σ$_{[+NEG]}$	Egyptian Arabic	(58)

In light of the picture painted by the data above, I close this chapter by asking the following question: are there any data that unequivocally show that featural clashes in English (or some other language) above the eventive core are well-formed, and thus an approach predicated on featural non-distinctness like my own cannot be on the right track?

Anand et al. (2021) present putative cases of this sort, arguing that they provide (i) counterevidence to my identity condition and (ii) support for approaches like Rudin 2019 and Chung 2013a that attribute a special status to the eventive core.[94] Specifically, in responding to the identity condition proposed here, Anand et al. highlight the following data. I have added the ellipsis sites for ease of exposition, while the putative antecedent is in square brackets:

(79) *Anand et al. 2021 (10)a-b*

a. She distanced herself from President Obama's heavily criticized comments about how ISIS is "contained," arguing that "it **cannot** be contained, [it **must** be defeated]." (How < *ellipsis site* > is still anyone's guess, in either party.)

b. This was a problem that [string theory **had to** solve], but for a long time it was

---

[94] These authors are responding to Ranero 2019b, an earlier version of the contents of this chapter that discusses examples like the Spanish (80).

not clear how <*ellipsis site*>.

Staring with the two data points above and what they are claimed to show, Anand et al. first observe that the putative antecedents in (79)a-b have circumstantial necessity modals. They then note that "*In the case of (10), the elided clause is understood as including a possibility modal like could or might, rather than the necessity modal of the apparent antecedent…*" (Anand et al. 2021: 5) This is taken to show that an approach like the one proposed here is on the wrong track (see below for more data). Unfortunately, Anand et al. fall into the methodological equivocation discussed in chapter 1 of this dissertation.

Let us take a step back and ponder briefly on the kind of data that would unequivocally demonstrate that the non-distinctness approach is on the wrong track—namely, well-formed examples in which the remnant is manipulated in order to force a content for the ellipsis site that clashes featurally with the antecedent, and there is no plausible content for the ellipsis site that would satisfy the non-distinctness requirement. I repeat some relevant Spanish data below in order to illustrate how a manipulation of the remnant can force a specific content for the ellipsis site. Given the adverbial remnant *en el futuro* 'in the future', there cannot be any tense other than future inside the ellipsis site; in this case, the example is ill-formed, as predicted by my approach:

(80)   *Spanish—tense clash forced by remnant (repeated from (23)*
    a. En el pasado, María **ha**   leído mucho y   Elena en el futuro
        in the past   María **have.PRES** read a.lot   and Elena in the future
        habrá   leído mucho también.
        have.FUT read a.lot   also
        'María has read a lot in the past, and in the future, Elena will have read a lot as well.'
    b. *En el   pasado, María **ha**   leído mucho y   Elena en el   futuro
         in the past   María **have.PRES** read a.lot   and Elena in the future
         <**habrá**   leído mucho> también.
          have.FUT read   a.lot   also
                                              (adapted from Murguia 2004:86 *apud* Saab 2016)

In contrast to the Spanish examples above, a single content for the ellipsis site is not forced in (79), such that the resulting judgement could inform the proper formulation of the identity condition on ellipsis. In fact, the authors themselves make this observation in passing, since they note that the ellipsis site in (79)a could take as antecedent *it cannot be contained*, which would give rise to a polarity mismatch (ruled-in by the condition proposed here)—see Anand et al.'s footnote 4. I take this to mean that the string in (79)a is available because there is, in fact, an ellipsis site that satisfies featural non-distinctness— it is irrelevant that there are alternative readings (that serve the same sort of communicative function) that would involve an ellipsis site that clashed featurally with the antecedent. Regarding (79)b, there also exists a possible content for the ellipsis site that satisfies featural non-distinctness—<to solve the problem>. This ellipsis site does not give rise to the putative modality clash that would argue against the identity condition in (77). Once more, it does not matter if consultants offer several possible interpretations for the examples in (79) and the interpretation that is offered most frequently for the ellipsis site violates featural non-distinctness, whereas other possible contents for the ellipsis site satisfying (77) (again, that serve the same sort of communicative function) are offered less frequently.

Consider now the data below, which are also used in Anand et al. 2021 as evidence against the condition defended here and in support of conditions akin to Rudin 2019's:

(81)   *Anand et al. 2020 (11)a-b*
      a. She tried to imagine his long, sensitive fingers cleaning and oiling a gun. She understood that [he must be surviving], but she couldn't figure out how <*ellipsis site*>.
      b. so he takes gravity instead as an 'endeavor,' capable of being certified 'mathematically' but itself now requiring a cause. But [this cause must somehow bridge the intervening spaces]. How <*ellipsis site*>?

The authors comment that in addition to interpretations where the ellipsis sites contain a possibility modal, "…*for the two cases in (11) interpretations in terms of a simple past (in (11)a) or simple present (in (11)b) are also **often available**.*" (Anand et al. 2021; emphasis mine). To reach this conclusion, the authors note that they polled 25 English speakers via e-mail and asked them to paraphrase the examples above. The consultants gave the range of possibilities described, with only one consultant giving the paraphrase where there was a full match for (81)b—i.e., <he must be surviving>.

As in the earlier cases discussed here, the authors have not provided evidence that there is a *single* content for the ellipsis sites in (81) that is forced by the remnant and results in a (well-formed) featural clash. Speakers accept contents for the ellipsis sites here that satisfy featural non-distinctness, even if such possibilities are dispreferred over conceivable alternatives containing modals. In particular, stressing the *wh*-remnant *how* in (81)a renders the fully matching content for the ellipsis site much more natural as a paraphrase, while a possible content for the ellipsis site in (81)b satisfies featural non-distinctness—<does this cause bridge the intervening spaces>. I conclude, thus, that Anand et al. 2021's data do not, in fact, challenge the conclusions I have drawn up to this point.

We are now ready to take the identity condition proposed thus far in this dissertation and apply it to a final empirical domain. In chapter 4, we will consider the nominal domain in several languages and assess a final group of mismatches—those involving [GENDER] and [NUMBER] features.

Chapter 4: Mismatches in the nominal domain

In this chapter, I discuss mismatches under ellipsis of features within the *nominal* domain, focusing specifically on [GENDER] and [NUMBER]. The data I will present show, in line with previous findings by Nunes & Zocca 2009, Bobaljik & Zocca 2011, Merchant 2014, Alexiadou 2017, Sudo & Spathas 2016, 2020a,b, Donatelli 2019, and Polinsky 2020, that the status of [GENDER] mismatches is not uniform. Instead, it is contingent on noun class membership: in a variety of languages, there exist three classes of nouns, where for each class, [GENDER] mismatches are well-formed or ill-formed in a manner that is distinct from the manner in which mismatches behave in the other classes.

The goal of this chapter is to show that the identity condition on ellipsis proposed thus far in this dissertation constitutes a key component for an account of the attested patterns in [GENDER] mismatches. In other words, the analysis that I will propose for the data here will be informed by the results from chapters 2 and 3, which led us to conclude that an identity condition predicated on the satisfaction of featural non-distinctness is on the right track. I will argue, however, that two other analytical ingredients (which have received prior support in the literature) are required in addition to the identity condition I have argued for thus far. First, I will propose that certain mismatches are well-formed due to repair-by-ellipsis of a morphophonological issue. Conversely, I will argue that certain mismatches that satisfy the identity condition proposed here are nevertheless ill-formed, because the configurations violate grammatical conditions that are not repairable by ellipsis (Mendes & Nevins 2021; see 4.3.3). Second, I will argue that the identity condition on ellipsis must be split into two statements. The first statement is the requirement that features satisfy non-distinctness. The other statement does not pertain to features, but √ROOTs.

Unlike the condition on features, however, √ROOTs in the antecedent and ellipsis site must be identical (see Saab 2008 and more details in section 4.4.2):

(1)   *Identity condition on √ROOTs*
        There must be a strict one-to-one match between all √ROOTs properly contained in the ellipsis site and √ROOTs in the antecedent.

As I will show, the combination of analytical ingredients discussed in this chapter allows for the right kind of flexibility to account for a significant degree of microvariation in the general pattern of [GENDER] mismatches that exists cross- and intra-linguistically (see 4.7).

This chapter is structured as follows. In section 4.1, I revisit the results of chapters 2 and 3 to remind the reader of the kind of identity condition that I concluded is necessary independently of the empirical generalizations that will be analyzed in this particular chapter. In other words, the goal of 4.1 is to delimit the scope of possible explanations that could be entertained for the [GENDER] mismatch data, given prior results from other empirical domains. Moving on to 4.2, I lay out the big-picture empirical generalizations that need to be explained. I show that there exist distinct classes of nouns under ellipsis that differ in their [GENDER] mismatch behavior: Class I nouns allow mismatches symmetrically, Class II nouns ban them symmetrically, and Class III nouns exhibit an asymmetric pattern, wherein only the masculine noun is possible in the antecedent (in cases of mismatch). I introduce data from several languages first in 4.2.1 but focus on Spanish in 4.2.2. In particular, I contrast the behavior of [GENDER] mismatches in opposition to [NUMBER]: whereas we observe an intricate pattern of possibilities in relation to [GENDER] mismatches, [NUMBER] seems to mismatch freely.

In 4.3, I set the stage for the proposal and discuss several analytical assumptions underlying my analysis. First, I discuss the derivation of NP ellipsis in 4.3.1, while in 4.3.2

I go into the structure of nominals, the locus of φ-features in the nominal spine, and the status of √ROOTs as syntactic primitives. Moving on to sub-section 4.3.3, I discuss repair-by-ellipsis, distinguishing between morphophonological gaps that can extraordinarily occur inside ellipsis sites, and other kinds of grammatically deviant configurations that cannot. In 4.3.4, I describe the difference between subject-predicate agreement and concord, presenting arguments that the latter is not mediated by the operation Agree.

With these theoretical assumptions in place, I delve into my analysis in 4.4., starting with [GENDER] mismatches with Class I and II nominals 4.4.1. In 4.4.2, I discuss Class III nominals and argue that, in order to derive the asymmetric mismatch pattern they exhibit, the identity condition proposed in this dissertation must be supplemented with a second statement that requires absolute √ROOT identity between antecedent and ellipsis site. I justify the requirement of √ROOT identity independently and show how the final identity condition I propose derives the Class III data.

In 4.5, I expand the empirical domain under analysis, showing how the proposal in the prior section can derive the behavior of [GENDER] mismatches with a larger sample of Spanish nouns. The derivation of an array of ill-formed mismatches in this section provides support for the proposal that certain lexical gaps are not repairable by ellipsis, whereas others are. Moving on to section 4.6, I briefly consider previous analyses for the core phenomenon and conclude the following: previous analyses either (i) cannot derive the totality of the pattern or (ii) they are challenged by the conclusions that were drawn in chapters 2 and 3 that delimited the scope of possible identity conditions. Section 4.7 lays out microvariation that has been reported across and within languages regarding [GENDER] mismatches and discusses how the analysis proposed here has the correct ingredients to

account for the data. In section 4.8, I return to the [NUMBER] feature and address a claim that number mismatches *are* ill-formed when *pluralia tantum* nouns are manipulated (Picallo 2017). I show that a closer look at the relevant data reveals that the problem with those mismatches is orthogonal to ellipsis. Finally, 4.9 concludes.

4.1 Recapitulating the results from previous chapters

Before delving into the specifics of our investigation into [GENDER] and [NUMBER] mismatches under ellipsis, let us recapitulate the conclusions of this dissertation so far. This exercise will allow us to reduce the space of analytical possibilities that should be pursuable *a priori* in order to explain the data analyzed in this chapter. In other words, the analysis proposed in what follows cannot exist in a vacuum that is dissociable from the conclusions drawn in previous chapters, where we focused on voice and argument structure mismatches (chapter 2) and cases where the mismatching feature is above the VoiceP layer (chapter 3).

First, let me emphasize once again that I reject the possibility that there are *different* identity conditions for different elliptical constructions. In other words, I continue to assume that constructions labelled sluicing, VP ellipsis, NP ellipsis, etc., instantiate the same phenomenon that is governed by the same identity condition. The only difference between these constructions lies on the head bearing the [E]-feature. For example, Spanish does not have aux-stranding VP ellipsis, whereas English does. We can straightforwardly encode this variation if VP ellipsis involves deletion of the complement of a $Voice^0$ head bearing the [E] feature (Merchant 2013a) and assuming that the English lexicon contains such a head, but the Spanish lexicon does not. My guiding principle, however, is that the identity condition in all languages, across all elliptical constructions, does *not* vary.

Given this assumption, let us remind ourselves of the conclusions drawn in chapter 2. In that chapter, I discussed English mismatches at the VoiceP/vP level first and showed that the unavailability of voice and argument structure mismatches in sluicing is strong evidence that there must be a syntactic component to the identity condition (Merchant 2013a)—i.e., an exclusively semantic identity condition fails to account for the ill-formed status of these data. In other words, syntax must be a core component of the identity condition. In this chapter, I will attempt to derive the patterns observed with [GENDER] under ellipsis solely via a syntactic condition. Note, however, that this does not mean that semantics (and even pragmatics) plays no role in deriving the data (see the discussion later in this chapter regarding encyclopedic gaps and how they cannot be repaired by ellipsis).

We then expanded the empirical picture and analyzed sluicing examples in Kaqchikel. We observed that a subset of voice mismatches is well-formed in that language, in contrast to English. I proposed as a result that *strict* identity of *features* cannot be the condition imposed in ellipsis. Instead, an identity condition predicated on featural non-distinctness delivers the correct results:

(2)   *Syntactic identity condition on ellipsis (first version; to be amended)*
      The antecedent and material properly contained in the ellipsis site must be featurally non-distinct.

I proposed that in all well-formed voice mismatches in Kaqchikel, either the antecedent or the ellipsis site lack Voice^0, complying with the condition in (2).

The main takeaway from chapter 2 that will inform the discussion ahead is that any account of the behavior of φ-feature mismatches within the nominal domain *cannot* resort to strict syntactic identity as a core component (more on this below). Nor can we resort to a condition requiring that the ellipsis site be a featural subset of the antecedent—recall my

analysis that in the Agent Focus–Active mismatches, the ellipsis site contains a Voice$_{ACT}$ that is absent in the antecedent (see chapter 2). As a consequence, I will propose an analysis where the well- or ill-formed status of φ-feature mismatches is predicated in part on the satisfaction of featural non-distinctness (2). In other words, the condition proposed in this dissertation predicts—all else being equal— that a featural *clash* will result in ill-formed ellipsis, whereas mismatches that do not involve a clash will result in well-formed ellipsis.

Moving on to chapter 3, the empirical domain assessed therein provided converging evidence for the identity condition in (2). I structured my argumentation as a counterpoint to the syntactic identity condition proposed in Rudin 2019, which posited as one of its components that only elements within the eventive core (i.e., the structure under VoiceP) enter into the calculation of identity. This aspect of the proposal in Rudin 2019 derived the asymmetry between mismatches above VoiceP (all seemingly well-formed in English) and mismatches within the eventive core (ill-formed in English). I showed, however, that this solution is not tenable, given that it fails to account for a broader sample of cross-linguistic data (e.g., Spanish data that clearly demonstrate that tense clashes are ill-formed), whereas the identity condition proposed here is better supported empirically. Overall, then, the discussion in chapter 3 led us to conclude that what regulates the well-formed or ill-formed status of mismatches under ellipsis is not the locus of the mismatching feature, but rather, the nature of the mismatch—if there is a featural *clash*, ellipsis is ill-formed.

Putting it all together, I will assume that a syntactic identity condition that ranges over the entirety of the ellipsis site and is predicated on featural non-distinctness constitutes one of the key components in any explanation of φ-feature mismatches in the nominal domain. We will see, however, that this condition needs to be coupled with additional ingredients

(albeit, ones that find independent support in the literature). With these assumptions in place, we can now turn to the empirical picture that forms the backbone of this chapter.

4.2 The data: [GENDER] and [NUMBER] mismatches

In this section, I discuss the broad empirical generalizations that have been established cross-linguistically regarding mismatches in [GENDER] and [NUMBER] under ellipsis (though I return briefly to [NUMBER] in 4.8). The core cross-linguistic pattern discussed in subsection 4.2.1 forms the basic puzzle that any approach to the identity condition on ellipsis needs to explain. I present Spanish data in 4.2.2 as a case study that illustrates the puzzle, and in 4.3 and 4.4, I go on to analyze those specific data.

The data explored in this section come from languages that encode grammatical gender and number distinctions on nominals, which will be analyzed using [GENDER] and [NUMBER] features in the syntax. For our purposes, I adopt the following definition of grammatical gender (Corbett 1991; see Kramer 2015: 65 for discussion):[95]

(3)   *Grammatical gender*
      Gender is the sorting of nouns into two or more classes as reflected in agreement
      morphology on determiners, verbs, and other syntactic categories.

When grammatical gender is manipulated under ellipsis, we will observe that similar empirical generalizations can be drawn from Greek, Brazilian Portuguese, Russian, and Spanish. In all these languages, some nouns come in pairs that are distinguished by their formal [GENDER] and [NUMBER] feature, resulting in singular and plural, as well as 'feminine' and 'masculine' variants of the same nominal. In Spanish, for example, the

---

[95] Kramer 2015 provides a more detailed final definition that incorporates the types of distinctions that languages make that fall under the umbrella of gender; e.g., animacy or human-ness distinctions.

[GENDER] feature distinguishes two versions of a noun, one meaning 'aunt' (*tía*) and the other 'uncle' (*tío*). The different featural representation of each version of the noun can be detected in this case via (i) the final vowel on the noun (the 'theme' vowel), which alternates between the two members of the noun pair, (ii) the choice of determiner that is compatible with each of the members of the noun pair, and (iii) concord on adjectival modifiers, which differs for each:

(4)   *Spanish nouns that alternate in [GENDER]: Class I*

    a. la   tí-a   generos-a de Bianca
       DET.F aunt-A generous-F of Bianca
       'Bianca's generous aunt'

    b. la-s   tí-a-s   generos-a-s   de Bianca
       DET.F-PL aunt-A-PL generous-F-PL of Bianca
       'Bianca's generous aunts'

    c. el   tí-o   generos-o de Bianca
       DET.M uncle-O generous-M of Bianca
       'Bianca's generous uncle'

    d. lo-s   tí-o-s   generos-o-s   de Bianca
       DET.M-PL uncle-O-PL generous-M-PL of Bianca
       'Bianca's generous uncles / uncles and aunts'

We will return briefly to the derivation of theme vowels in 4.5.1.

Consider now noun pairs like *abogada/abogado* 'female lawyer/male lawyer', which are parallel to the example above in the morphological alternations that they participate in:

(5)   *Spanish nouns that alternate in [GENDER]: Class II*

    a. la   abogad-a exitos-a
       DET.F lawyer-A successful-F
       'the successful (female) lawyer'

    b. la-s   abogad-a-s exitos-a-s
       DET.F-PL lawyer-A-PL successful-F-PL
       'the successful (female) lawyers'

    c. el   abogad-o exitos-o
       DET.M lawyer-O successful-M
       'the successful (male) lawyer'

d. lo-s   abogad-o-s exitos-o-s
   DET.M-PL lawyer-O-PL successful-M-PL
   'the successful (male) lawyers / lawyers'

 As we will see in section 4.2.2, the behavior of this noun pair in ellipsis is different from the behavior of the noun pair in (4).

    Finally, the behavior under ellipsis of yet another class of noun pair is different both from the behavior of pairs like (4) and of pairs like (5). In this third class, the feminine version is distinguishable on the surface from the masculine version because it bears an additional suffix. Observe, for example, the noun pair *actriz/actor* 'actress/actor', where the feminine version bears the *-iz* suffix (note that orthographically, the suffix is written as <ic> in the plural):

(6)   *Spanish nouns that alternate in [GENDER]: III*

    a. la   actr-**iz** invitad-a
      DET.F actor-**IZ** invited-F
      'the invited actress'

    b. la-s   actr-**ic**-es   invitad-a-s
      DET.F-PL actor-**IZ**-PL invited-F-PL
      'the invited actresses'

    c. el   actor invitad-o
      DET.M actor invited-M
       'the invited actor'

    d. lo-s   actor-es invitad-o-s
      DET.M-PL actor-PL invited-M-PL
       'the invited actors / actors and actresses'

The empirical picture that will arise regarding ellipsis and mismatches in [GENDER] is complex. However, we will observe that there exists a *general* pattern cross-linguistically, wherein nouns belonging to different classes of nouns exhibit different mismatch behavior under ellipsis. This general pattern has been reported and analyzed for several languages, with some variation regarding the scope of the empirical description available for each language and the degree to which analyses rely on syntactic or semantic components of the

identity condition to account for the data (see section 4.7 on microvariation; Nunes &
Zocca 2009, Bobaljik & Zocca 2011 for Brazilian Portuguese; Merchant 2014, Alexiadou
2017, Sudo & Spathas 2016, 2020a,b for Greek; Depiante & Masullo 2001, Saab 2004,
2008, 2010a, Picallo 2017, Donatelli 2019 for Spanish; Polinsky 2020 for Russian). Let us
now turn to the data.

4.2.1 Cross-linguistic generalizations

The broad generalization on [GENDER] mismatches under ellipsis is the following: there
exist distinct classes of nouns, where for each class, the mismatch behavior is distinct. We
find that there are three such classes of nouns in several languages:

(7)   *[GENDER] mismatches under ellipsis—three classes of nouns*
    a. *Class I*:   [GENDER] mismatch symmetrically ill-formed   (8), (11), (14), (17)
    b. *Class II*: [GENDER] mismatch symmetrically well-formed (9), (12), (15), (18)
    c. *Class III:* [GENDER] mismatch asymmetrically well-formed   (10), (16), (19)
        (i)   Masculine antecedent ✓
        (ii)   Feminine antecedent   *

The generalization in (7) can be illustrated through data from different languages: Brazilian
Portuguese, Greek, and Russian.[96, 97]

---

[96] Bobaljik & Zocca 2011 provide some German judgments as well. The paper also lists noun pairs in German
and Romanian that fall into the class in which [GENDER] mismatches are symmetrically ill-formed, as well
as examples of noun pairs that fall into the class in which [GENDER] mismatches display asymmetrical
behavior. The ellipsis examples are not given, though. For discussion of other languages, see Murphy et al.
2018 (Bosnian-Croatian-Serbian) and Barrie 2016 (Cayuga).

[97] English could also provide a testing ground for mismatches in grammatical gender, since there exist noun
pairs like *waiter/waitress* that encode such a distinction, at least for some speakers (see Sprouse et al. 2020).
Looking at pronouns as well, Sag 1976: 123 (see endnote 15) reports that sloppy identity readings are difficult
to access for some English speakers when reflexives mismatch in [GENDER]. In other words, it seems that a

First, let us consider predicate ellipsis in Brazilian Portuguese. In these examples, a mismatch is forced via the manipulation of a proper name in the remnant. The assumption is that a mismatch is forced because the individuals under discussion use masculine or feminine pronouns. Notice that in (8)a-b, a mismatch is symmetrically *ill*-formed: neither the masculine nor feminine version of the noun can be in the antecedent. Examples (8)-(10) are adapted from Bobaljik & Zocca 2011:[98]

(8)   *[GENDER] mismatch symmetrically ill-formed; Brazilian Portuguese*
    a. # O   Zé vai ser **ti-o**   e   a   Lu também vai ser <**ti-a**>.
        the Zé will be **uncle-M** and the Lu also   will be   **aunt-F**
        Intended: 'Zé will become an uncle and Lu will too.'
    b. #A Lu vai ser **ti-a**   e   o   Zé também vai ser <**ti-o**>.
        the Lu will be **aunt-F** and the Zé also   will be   **uncle-M**
        Intended: 'Lu will become an aunt and Lu will too.'

In contrast, a mismatch is symmetrically *well*-formed in the examples below: either the masculine (9)a or feminine (9)b version of the noun pair can be in the antecedent.

---

sloppy interpretation is accessible for all speakers in an example like (i)a, whereas it is less accessible for a subset of speakers in an example like (i)b. The mismatch in (i)b is forced by the manipulation of a proper name in the remnant of an individual who is known to use masculine pronouns.

(i)   *Sloppy identity and gender mismatches in VPE (based on Sag 1976)*
    a.   Bianca$_i$ scratched her$_i$ arm, and Raja$_j$ did <scratch her$_j$ arm> too.
    b.   Bianca$_i$ scratched her$_i$ arm, and Ben$_j$ did <scratch her$_i$/??his$_j$ arm> too.

One way to interpret the above is that a sloppy reading in (i)b creates a featural clash for some, thus violating the identity condition proposed in this dissertation. It would be interesting to know how speakers today whose grammars differ in their featural representation of pronouns (including allowing specific, singular *they*; see Conrod 2019, Konnelly & Cowper 2020) judge different manipulations that force sloppy interpretations.

[98] Bobaljik & Zocca use # to notate ill-formedness and do not provide the baseline examples showing that all of these examples are well-formed in their non-elliptical versions. From the prose, however, it is safe to assume that the pattern observed is about ellipsis specifically.

(9)   *[GENDER] mismatch symmetrically well-formed; Brazilian Portuguese*

    a. O Pedro é **medic-o** e   a Marta também é <**médic-a**>.
       the Pedro is **doctor-M** and the Marta also   is   **doctor-F**
       'Pedro is a doctor and Marta is too.'

    b. A   Marta é **medic-a** e   o Pedro também é <**medic-o**>.
       the Marta is **doctor-F** and the Pedro also   is   **doctor-M**
       'Marta is a doctor and Pedro is too.'

Finally, the examples in (10)a-b from Bobaljik & Zocca 2011 illustrate an asymmetrical mismatch behavior—whereas the masculine version of the noun pair is well-formed in the antecedent (10)a, the feminine version is ill-formed in the antecedent (10)b:

(10)   *[GENDER] mismatch asymmetrically well-formed; Brazilian Portuguese*

    a. ?O   Paulo é **ator** e   a   Fernanda também é <**atr-iz**>.
       the Paulo is **actor** and the Fernanda also   is   **actr-IZ**
      'Paulo is an actor and Fernanda is too.'

    b. #A   Fernanda é **atr-iz** e   o   Paulo também é <**ator**>.
       the Fernanda is **actr-IZ** and the Paulo also   is   **actor**
      Intended: 'Fernanda is an actress and Paulo is too.'

The contrast between noun pairs like in (8) and (9) is not specific to predicate ellipsis; the same contrast is maintained under NP ellipsis, as the following examples adapted from Nunes & Zocca 2009 show—note that the non-elliptical controls are not provided by the authors, but the prose in the paper makes clear that the ill-formed status of (11) is due to ellipsis. The way to force the mismatch here is via a remnant determiner which bears a morphological reflex of the [GENDER] specification of the nominal in the ellipsis site:

(11)   *[GENDER] mismatch ill-formed; Brazilian Portuguese*

    *O João visitou o   **tio**   dele e   o   Pedro prometeu visitar **a**   dele.
       the João visited the.M **uncle** of.his and the Pedro promised visit   **the.F** of.his
       Intended: 'João visited his uncle and Pedro promised to visit his aunt.'

(12)   *[GENDER] mismatch well-formed; Brazilian Portuguese*

    O João visitou o   **médico** dele e   o   Pedro visitou **a**   dele.
      the João visited the.M **doctor.M** of.his and the Pedro visited **the.F** of.his
      'João visited his doctor and Pedro visited his (female) doctor.'

226

While Nunes & Zocca do not provide examples with the third class of nouns, I complete

the paradigm with the examples below (Gesoel Mendes and Jessica Mendes p.c.):

(13)   *[GENDER] mismatch asymmetrically ill-formed; Brazilian Portuguese*

    a. O   João visitou o   **ator**   da Globo e   a   Maria visitou a   **atriz**
       the João visited the **actor** of Globo and the Maria visited the.F **actress**
       da Band.
       of Band
       'João visited the actor from Globo and Maria visited the actress from Band.'

    b. ?O João visitou o **ator** da Globo e a Maria visitou a <**atriz**> da Band.

    c. O   João visitou a   **atriz**   da Globo e   a   Maria visitou o   **ator**
       the João visited the **actress** of Globo and the Maria visited the.M **actor**
       da Band.
       of Band
       'João visited the actress from Globo and Maria visited the actor from Band.'

    d. *O João visitou a **atriz** da Globo e a Maria visitou o <**ator**> da Band.

The same general pattern can be illustrated through NP ellipsis in Greek (Sudo & 2016;

data adapted from Alexiadou 2017). In these examples as well, the mismatch is forced by

manipulating elements in the remnant that display morphology that reflects the [GENDER]

specification of the nominal in the ellipsis site; e.g., the remnant *mia* 'one' in (14)a encodes

feminine features, whereas the remnant *enan* 'one' in (14)b encodes masculine features:[99]

---

[99] Merchant (2014) also provides examples of predicate ellipsis illustrating the same three-way asymmetry.

However, Merchant argues that mismatches under NP ellipsis are all ungrammatical (see his examples (10),

(23), (26)), irrespective of the class of nouns (see similar comments for Spanish in van Craenenbroeck &

Merchant 2013: 736-737 and section 4.7 below). Sudo & Spathas 2016 show that Merchant's NP ellipsis

examples are ill-formed for reasons independent of ellipsis, so the presumed contrast between predicate and

NP ellipsis that Merchant 2014 sought to derive is not supported empirically. Alexiadou (2017) seems to

agree with Sudo & Spathas' response to Merchant 2014, since she reports the empirical correction made by

these authors and provides an analysis. In a later paper, Sudo & Spathas 2020a report that they consulted 8

naïve Greek consultants and 16 native speaker linguists and found that the three-way asymmetry in predicate

(14)   *[GENDER] mismatch symmetrically ill-formed; Greek*

    a. *O Petros episkefthike enan adherfo tu sti   Veria, ke mia.
       the Petros visited   one.M brother his in.the Veria and one.F
      stin   Katerini
      in.the Katerini
      Intended: 'Petros visited a brother of his in Veria and a (sister) in Katerini.'

    b. *O Petros episkefthike mia adherfi tu sti   Veria, ke enan
       the Petros visited   one.F sister   his in.the Veria and one.M
      stin   Katerini.
      in.the Katerini
      Intended: 'Petros visited a sister of his in Veria, and a (brother) in Katerini.'

(15)   *[GENDER] mismatch symmetrically well-formed; Greek*

    a. O Petros episkefthike ena   jatro sti   Veria, ke mia sti.   Katerini.
      the Petros visited   one.M doctor in.the Veria and one.F in.the Katerini
      'Petros visited a male doctor in Veria, and a female doctor in Katerini.'

    b. O Petros episkefthike mia   jatro sti Veria, ke enan sti Katerini.
      the Petros visited   one.F doctor in Veria and one.M in Katerini
      'Petros visited a female doctor in Veria, and a male doctor in Katerini.'

(16)   *[GENDER] mismatch asymmetrically well-formed; Greek*

    a. O Petros episkefthike enan dhaskalo sti   Veria, ke mia stin Katerini.
      the Petros visited   one.M teacher.M in.the Veria and one.F in.the Katerini
      'Petros visited a male teacher of his in Veria, and a female teacher in Katerini.'

    b. *O Petros episkefthike mia dhaskala sti   Veria, ke enan stin Katerini.
      the Petros visited   one.F teacher.F in.the Veria and one.M in.the Katerini
      Intended: 'Petros visited a female teacher of his in Veria, and a male teacher in Katerini.'

We observe, then, that there exists a three-way asymmetry in the availability of [GENDER]

mismatches in Greek as well.

---

ellipsis reported by Merchant 2014 is hard to replicate: while the Class I and II patterns seem fairly replicable,

the Class III contrast is not found uniformly, and whenever speakers do judge there to be an asymmetric

pattern with such a class, the effect is subtle. Furthermore, even though they showed in Sudo & Spathas 2016

that [GENDER] mismatches in NP ellipsis are also possible, Sudo & Spathas 2020a report that Class III

judgements are unstable across speakers regardless of ellipsis type (see Sudo & Spathas 2020a 8-12). See

section 4.7 for the issue of microvariation.

To end our illustration of the general pattern, Polinsky 2020 presents the results of a large survey of over one hundred Russian speakers and finds that the same three-way asymmetry exists in that language (Bobaljik & Zocca 2011 provided some Russian examples as well; see 4.7 below for discussion of microvariation). Examples of the three different classes of nouns in Russian predicate ellipsis are given below, adapted from Polinsky 2020 (except (19)a; Maria Polinsky p.c.):

(17)   *[GENDER] mismatch symmetrically ill-formed; Russian*
    a. *Orlov byl graf, i Rostova tože.
      Orlov was count and Rostova too
      Intended: 'Orlov was a count and Rostova was a countess.'
    b. *Rostova byla grafinja, i Lenin tože.
      Rostova was countess and Lenin too
      Intended: 'Rostova was a countess and Lenin too.'

(18)   *[GENDER] mismatch symmetrically well-formed; Russian*
    a. Petja dežurnyj, i Maša tože.
      Petja on.duty.M and Masha too
      'Petja is an on-duty person and Masha too.'
    b. Maša dežurnaja, i Petja tože.
      Masha on.duty.F and Petja too
      'Masha is an on-duty person and Petja too.'

(19)   *[GENDER] mismatch asymmetrically well-formed; Russian*
    a. Petja učitel', i Maša tože.
      Petja teacher and Masha too
      'Petja is a teacher and Masha too.'
    b. *Maša učitel'-nica, i Petja tože.
      Masha teacher-F and Petja too
      Intended: 'Masha is a teacher and Petja too.'

To summarize this subsection, we have observed that a similar pattern recurs cross-linguistically regarding [GENDER] mismatches under ellipsis. In a nutshell, not all noun pairs behave alike, leading us to posit the existence of three distinct classes of nouns in

each language. In the next subsection, I focus on Spanish to carve out a set of data which I will then analyze in section 4.4.

4.2.2 Spanish

In this section, I lay out the pattern of [GENDER] mismatches in Spanish predicate and NP ellipsis, presenting my own judgments and those of three other native speakers of Guatemalan Spanish. In section 4.7, I will revisit prior descriptions of Spanish and note the pockets of microvariation that seem to exist for the language (as well as cross-linguistically).

First, let us note a contrast between the possibility of mismatch in [NUMBER] versus [GENDER] features (a contrast that I will confirm is absolute in section 4.8). Just like in English, [NUMBER] mismatches in ellipsis are well-formed in Spanish; note that in the NP ellipsis examples in (21) below, the well-formed nature of the [NUMBER] mismatch is symmetrical. I am not aware of the existence of any class of nouns in Spanish where the pattern is different (see 4.8 below for a discussion of *pluralia tantum* nouns):

(20)   *[NUMBER] mismatches under ellipsis: English*
          a. Raven bought one cookie, but Raja bought three <cookies>.   ✓SG-PL
          b. Raven bought three cookies, but Raja bought only one <cookie>.   ✓PL-SG

(21)   *[NUMBER] mismatches under ellipsis: Spanish*
          a. El   gato de Gaby y   los   gato-s de Sebas son dócil-es.
              the.M cat.M of Gaby and the.M-PL cat.M-PL of Sebas are docile-PL
              'Gaby's (male) cat and Sebas's cats are docile.'
          b. El gato de Gaby y los <gatos> de Sebas son dóciles.   ✓SG-PL
          c. Lo-s   gato-s de Sebas y   el   gato de Gaby son dócil-es.
              the.M-PL cat-PL of Sebas and the.M cat.M of Gaby   are docile-PL
              'Sebas's cats and Gaby's (male) cat are docile.'
          d. Los gatos de Sebas y el <gato> de Gaby son dóciles.   ✓PL-SG

The mismatches under ellipsis in (21)b,d are as well-formed as the non-elliptical controls.

With this in mind, let us move on to mismatches in the grammatical [GENDER] feature (the literature on grammatical gender in Spanish is vast; the reader is referred to Roca 1989, Harris 1991, Aronoff 1994, Alexiadou 2004, Kramer 2015, Picallo 2008, 2017). In Spanish, a subset of nouns come in pairs, where one member of the pair has traditionally been called 'masculine' and the other 'feminine'; I take this distinction to be represented in the syntax via [GENDER] features, which trigger detectable reflexes on the surface (such as concord on determiners and nominal modifiers). For example, the determiners *la/el* alternate depending on the nominal they associate with, as does the concord on the adjectival modifier *exitosa/o* 'successful'. Notice as well that certain modifiers like *independiente* 'independent' are invariant and do not exhibit an overt reflex of concord at all:

(22)   *[GENDER] and its morphological reflexes*
        a. la   catedrática exitos-a   e   independiente
           the.F professor.F successful-F and independent
           'the successful and independent (female) university professor'
        b. el   catedrático exitos-o   e   independiente
           the.M professor.M successful-M and independent
           'the successful and independent (male) university professor'

We will return to the syntactic mechanism underlying concord in section 4.3.4.

The initial picture that arises regarding ellipsis and [GENDER] mismatches for Spanish is the following, where three different patterns arise (much like our discussion in the previous subsection). Unfortunately, authors have been inconsistent in their labelling of

which class of nouns exemplifies which pattern. I have chosen to use Merchant 2014's taxonomy:[100]

(23)   *[GENDER] mismatches under ellipsis—three classes of nouns*
        a. *Class I*: [GENDER] mismatch symmetrically ill-formed   e.g. *tío ~ tía*
        b. *Class II*: [GENDER] mismatch symmetrically well-formed
                                                                            e.g. *abogado ~ abogada*
        c. *Class III:* [GENDER] mismatch asymmetrically well-formed   e.g. *actor ~ actriz*
            (i)   Masculine antecedent ✓
            (ii)   Feminine antecedent *

I will illustrate each of the classes of nouns using predicate and NP ellipsis. At this juncture,

it is important to note that it is the status of judgments *relative to each other* that is crucial,

rather than the *absolute* status of judgments on sentences in isolation (see Merchant 2014

fn. 6 for similar comments and section 4.9 here). For instance, while none of the ellipsis

examples with Class III nouns are perfect, there is a clear contrast within that class when

one compares the example where the masculine version of the noun is in the antecedent,

versus the examples where the feminine version of the noun is in the antecedent. Even

though the former mismatch is not perfect (masculine-feminine), the latter (feminine-

masculine) is simply impossible.

---

[100] Merchant 2014 reports predicate ellipsis examples in Spanish from Depiante & Masullo 2001 showing this asymmetry. The same overall picture is also reported by Donatelli 2019, with two caveats. The first is purely notational—Donatelli follows Bobaljik & Zocca 2011 in labeling Class I nouns those that allow a mismatch symmetrically (i.e., our Class II); Class II those that allow a mismatch asymmetrically (i.e., our Class III); and Class III those that disallow a mismatch symmetrically (i.e., our Class I). Second, Donatelli 2019: chapter 4 only reports predicate ellipsis examples, except its data in (11)-(13) which involve NP ellipsis (citing Kornfeld & Saab 2004 and Saab 2010a). Those NP ellipsis examples involve ill-formed mismatches with my Class I (*tío/tía* 'uncle/aunt'), and Class IV and Class V nouns (*padre/madre* 'father/mother'; see 4.5.1).

Let us begin with Class I. Some nouns in this class are given below (note that all lists I present henceforth are not exhaustive, and there appears to be microvariation regarding membership into each class of nouns for different speaker populations). Notice that the masculine and feminine versions of nouns in this class clearly share the same root:[101]

(24)　　*Class I nouns*
　　　　a.　　*Kinship terms*
　　　　　　tío/tía　　　　　　　'uncle/aunt'
　　　　　　primo/prima　　　　'cousin'
　　　　　　hermano/hermana　'brother/sister'
　　　　　　hijo/hija　　　　　　'son/daughter'
　　　　　　cuñado/cuñada　　　'brother-in-law/sister-in-law'
　　　　　　suegro/suegra　　　'father-in-law/mother-in-law'
　　　　b.　　*Animals*
　　　　　　perro/perra　　　　'dog'
　　　　　　gato/gata　　　　　'cat'
　　　　　　conejo/coneja　　　'rabbit'
　　　　　　cerdo/cerda　　　　'pig'

Nouns in this class include kinship terms and animals (see Saab 2008: 506, Picallo 2017:6 for other judgements using nouns denoting animals). Consider the noun pair *gato/gata* 'cat': the masculine bears the theme vowel *-o*, whereas the feminine bears *-a*:

(25)　　El　　gato　de　Gaby　es　dócil.
　　　　the.M cat.M of　Gaby　is　docile
　　　　'Gaby's (male) cat is docile.'

(26)　　La　　gata　de　Gaby　es　dócil.
　　　　the.F cat.M of　Gaby　is　docile
　　　　'Gaby's (female) cat is docile.'

In section 4.4.1, I will briefly lay out how theme vowels come about, following Kramer 2015's analysis that they instantiate the spell-out of a node that is inserted in the post-

---

[101] Most authors have also placed nobility terms in this Class; e.g. *duque/duquesa* 'duke/duchess'. I set these aside until 4.7, since my consultants' judgments are not uniform.

syntactic component on the basis of arbitrary declension classes. For now, I point out that theme vowel alternations are not predictive of membership into a specific class of nouns (e.g., *tío/tía* 'uncle/aunt' is Class I, while *abogado/a* 'male/female lawyer' is Class II, as shown later on; see Bobaljik & Zocca 2011 for this issue in Brazilian Portuguese as well).

Turning to ellipsis, Class I nouns cannot mismatch in [GENDER]: i.e., there cannot be a masculine noun in the antecedent and its feminine counterpart in the ellipsis site or vice-versa. Representative NP ellipsis examples are shown below. Notice that a mismatch is forced in the ellipsis site by manipulating the remnant determiner in the target clause. In other words, we know that there must be a mismatch because of a morphological reflex outside the ellipsis site of the [GENDER] feature borne by the nominal inside the ellipsis site:

(27)　　*Class I and NPE; masculine-feminine*
　　　a. El　　gato　de Juan es dócil,　pero el　　gato　de María　es　feroz.
　　　　the.M cat.M of Juan　is docile, but　the.M cat.M of María　is　ferocious
　　　　'Juan's (male) cat is docile, but María's (male) cat is ferocious.'
　　　b. El gato de Juan es dócil, pero el <gato> de María es feroz.
　　　　'Juan's (male) cat is docile, but María's (male cat) is ferocious.'
　　　c. El　　gato　de Juan es dócil,　pero la　　gata　de María es feroz.
　　　　the.M cat.M of Juan　is　docile, but　the.F cat.M of　María is ferocious
　　　　'Juan's (male) cat is docile, but María's (female) cat is ferocious.'
　　　d. *El gato de Juan es dócil, pero la <gata> de María es feroz.
　　　　Intended: 'Juan's (male) cat is docile, but María's (female cat) is ferocious.'

The pattern is symmetrical; having a feminine noun in the antecedent is also ill-formed.

(28)　　*Class I and NPE: feminine-masculine*
　　　a. La　　gata de María es dócil, pero la　　gata　de Juan es feroz.
　　　　the.F cat.F of María is docile but　the.F. cat.F of　Juan is　ferocious
　　　　'María's (female) cat is docile, but Juan's (female) cat is ferocious.'
　　　b. La gata de María es dócil, pero la <gata> de Juan es feroz.
　　　　'María's (female) cat is docile, but Juan's (female cat) is ferocious.'

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

d. *La gata de María es dócil, pero el <gato> de Juan es feroz.
   Intended: 'María's (female) cat is docile, but Juan's (male cat) is ferocious.'

The pattern here is not an artifact of NP ellipsis specifically, since other elliptical constructions show the same results. Consider predicate ellipsis examples with a remnant *también* 'too' (see chapter 3). In this case, the entire nominal bearing the mismatching feature is elided. Just like in the NP ellipsis cases before, a [GENDER] mismatch is ill-formed with nouns in Class I. Notice that we ensure a match or mismatch in the ellipsis site through the remnant in the target clause. In this case, a pet's name is used to force the mismatch, where a prior context is presented laying out the sex of each cat:

(29)   *Class I and predicate ellipsis: masculine-feminine*
         Context: Ru is a cat person who has four cats, two male ones called Max and Lucas, and two female ones called Lisa and Lexie. We are describing Ru's cats.

a. Max es un gato   dócil, y   Lucas también es un   gato dócil.
   Max is a.M cat.M docile and Lucas also   is a.M cat.M docile
   'Max is a docile (male) cat, and Lucas is a docile (male) cat too.'

b. Max es un gato dócil, y Lucas también <es un gato dócil>.
   'Max is a docile (male) cat and Lucas is too.'

c. Max es un gato   dócil, y   Lisa también es una gata dócil.
   Max is a.M cat.M docile and Lisa also   is a.F cat.F docile
   'Max is a docile (male) cat, and Lisa is a docile (female) cat too.'

d. *Max es un gato dócil, y Lisa también <es una gata dócil>.
   Intended: 'Max is a docile (male) cat and Lisa is too.'

Just like with NPE, the effect is symmetrical—a feminine noun in the antecedent cannot mismatch with its masculine counterpart in the ellipsis site:

(30)   *Class I and predicate ellipsis: feminine-masculine*

a. Lisa es una gata dócil, y   Max también es un gato   dócil.
   Lisa is a.F cat.F docile and Max also   is a.M cat.M docile
   'Lisa is a docile (female) cat and Max is a docile (male) cat too.'

b. *Lisa es una gata dócil, y Max también <es un gato dócil>.
   Intended: 'Lisa is a docile (female) cat and Max is too.'

235

Other nouns in this class include kinship nouns like *tío/a* 'uncle/aunt'. Their behavior is identical. As noted before, we force the mismatch in predicate ellipsis via the manipulation of proper names associated by the speaker with individuals who use masculine or feminine pronouns:

(31)   *Class I and NP ellipsis*

    a. El   tío   de Ana y   el   tío   de Clara llegaron.
       the.M uncle of Ana and the.M uncle of Clara arrived
       'Ana's uncle and Clara's uncle arrived.'

    b. El tío de Ana y el <tío> de Clara llegaron.
       'Ana's uncle and Clara's (uncle) arrived.'

    c. El   tío   de Ana y   la   tía   de Clara llegaron.
       the.M uncle of Ana and the.F aunt of Clara arrived
       'Ana's uncle and Clara's aunt arrived.'

    d. *El tío de Ana y la <tía> de Clara llegaron.
       Intended: 'Ana's uncle and Clara's (aunt) arrived.'

    e. La   tía   de Ana y   la   tía de Clara llegaron.
       the.F aunt of Ana and the.F aunt of Clara arrived
       'Ana's aunt and Clara's's aunt arrived.'

    f. La tía de Ana y la <tía> de Clara llegaron.
       'Ana's aunt and Clara's (aunt) arrived.'

    g. La   tía   de Ana y   el   tío   de Clara llegaron.
       the.F aunt of Ana and the.M uncle of Clara arrived
       'Ana's aunt and Clara's uncle arrived.'

    h. *La tía de Ana y el <tío> de Clara llegaron.
       Intended: 'Ana's aunt and Clara's (uncle) arrived.'

(32)   *Class I and predicate ellipsis*

    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 already is uncle and Pablo also   already 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>.
       'Sebas is an uncle now, and Pablo is too.'

    c. Sebas ya   es tío, y   Gaby también ya   es tía.
       Sebas already is uncle and Gaby also   already 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>.
       Intended: 'Sebas is an uncle now, and Gaby is (an aunt now) too.'

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

    f. Gaby ya es tía, y Laura también <ya es tía>.
       'Gaby is an aunt now and Laura is too.'

    g. Gaby ya   es tía, y   Pablo también ya   es tío.
       Gaby already is aunt and Pablo also   already is uncle
       'Gaby is an aunt now and Pablo is an uncle now too.'

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

To summarize the data so far, Class I nouns disallow a [GENDER] mismatch symmetrically.

    Let us move on now to Class II. Notice that nouns in this class denote professions or occupations:

(33)   *Class II nouns*
       a. abogado/abogada   'lawyer'
       b. ingeniero/ingeniera   'engineer'
       c. arqueólogo/arqueóloga   'archeologist'
       d. ardinero/jardinera   'gardener'
       e. escritor/escritora   'writer'
       f. escultor/escultora   'sculptor'
       g. pintor/pintora   'painter'
       h. (el) testigo/(la) testigo   'witness'
       i. (el) artista/(la) artista   'artist'
       j. (el) lingüista/(la) lingüista 'linguist'

Just like with Class I nouns, the pairs in Class II clearly share the same root. In some of these noun pairs, the masculine bears the theme vowel *-o* and the feminine the theme vowel *-a* (e.g., *jardinero/a* 'gardener'). These nouns share the same morphological marking, then, as Class I nouns (e.g. *tío/tía* 'uncle/aunt'). In other Class II pairs, the masculine does not bear a theme vowel (e.g., *escultor* 'sculptor'). Finally, in a subset of pairs, the difference in [GENDER] specification between the masculine and feminine is only detectable via concord on determiners or adjectives (e.g., *(el) testigo* / *(la) testigo* 'witness'); I will return to these nouns (so-called *epicene* nouns) in section 4.7.

The behavior of Class II nouns under ellipsis is the diametrical opposite of the behavior of Class I nouns: [GENDER] mismatches in this case are symmetrically *well*-formed: having a masculine noun in the antecedent mismatch with a feminine noun in the ellipsis site is well-formed and vice-versa. Consider first the NP ellipsis examples below, which show that a masculine noun is allowed in the antecedent. The manipulation forcing the mismatch here is the determiner in the remnant:

(34)   *Class II and NPE: masculine-feminine*
    a. El   abogado de Ana es competente y   el   abogado de Eu también.
       the.M lawyer.M of Ana is competent and the.M lawyer.M of Eu too
       'Ana's (male) lawyer is competent and Eu's (male) lawyer is too.'
    b. El abogado de Ana es competente y el <abogado> de Eu también.
       'Ana's (male) lawyer is competent and Eu's (male lawyer) is too.'
    c. 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.'
    d. El abogado de Ana es competente y la <abogada> de Eu también.
       'Ana's (male) lawyer is competent and Eu's (female lawyer) is too.'

The examples above stand in stark contrast to the examples involving Class I nouns that were discussed previously and the pattern observed is parallel to the Brazilian Portuguese examples in (12) and the Greek examples in (15)a. Whereas any mismatch with Class I nouns is simply impossible, parallel examples using Class II nouns are acceptable.

Consider now the examples below—parallel to the Greek (15)b—which show that a mismatch is well-formed symmetrically: a feminine noun in the antecedent can also mismatch with its masculine counterpart in the ellipsis site:

(35)   *Class II and NPE: feminine-masculine*
    a. La   abogada de Ana es competente y   la   abogada de Eu también.
       the.F lawyer.F of Ana is competent and the.F lawyer.F of Eu too
       'Ana's (female) lawyer is competent and Eu's (female) lawyer is too.'
    b. La abogada de Ana es competente y la <abogada> de Eu también.
       'Ana's (female) 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.
   'Ana's (female) lawyer is competent and Eu's (male lawyer) is too.'

My consultants and I judge (35)d to be slightly degraded in comparison to (34)d, but the example is still acceptable and the contrast with all Class I mismatches is sharp.

Observe as well that, in a similar fashion to Class I nouns, the behavior of Class II nouns is not an artifact of NP ellipsis specifically. The symmetrical well-formedness of a [GENDER] mismatch is maintained under predicate ellipsis, where the entire nominal is elided. I illustrate here with examples where a polarity particle is in the remnant. Consider first a masculine noun in the antecedent and a feminine noun in the ellipsis site. Like before, a mismatch is forced by manipulating an element in the remnant of the target clause; in this case, once more, proper names that are associated by the speaker with individuals who use male or female pronouns:

(36)   *Class II and predicate ellipsis: masculine-feminine*
    a. Matías no es abogado, pero Tomás sí   es abogado.
       Matías not is lawyer.M but   Tomás yes is lawyer.M
       'Matías is not a (male) lawyer, but Tomás is a (male) lawyer.'

    b. Matías no es abogado, pero Tomás sí <es abogado>.
       'Matías is not a (male) lawyer, but Tomás is (a male lawyer).'

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

    d. Matías no es abogado, pero Rosa sí <es abogada>.
       'Matías is not a (male) lawyer, but Rosa is (a female lawyer).'

We observe the same pattern if the feminine noun is in the antecedent (though see section 4.7, where we discuss Saab 2004 and Donatelli 2019's comments on this particular manipulation):

(37) *Class II and predicate ellipsis: feminine-masculine*

  a. Rosa no es abogada, pero Mina sí <es abogada>.
   Rosa not is lawyer.F but Mina yes is alwyer.F
   'Rosa is not a lawyer, but Mina is a lawyer.'

  b. Rosa no es abogada, pero Mina sí <es abogada>.
   'Rosa is not a lawyer, but Mina is (a lawyer).'

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

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

To summarize the data so far, [GENDER] mismatches with Class II nouns are symmetrically well-formed in different elliptical constructions.

Finally, let us discuss Class III nouns. With nouns in this class, [GENDER] is allowed to mismatch under ellipsis, but in an asymmetrical fashion—only the masculine counterpart is allowed in the antecedent.[102]

---

[102] Saab 2010a and Donatelli 2019 discuss some noun pairs that they report ban mismatches altogether that a subset of my consultants would list as Class III. These noun pairs seem to be nobility terms that share the same root; e.g., *duque/duquesa* 'duke/duchess' (as discussed by Saab and Donatelli), but others like *marqués/marquesa* 'marquis/marquess', *emperador/emperatriz* 'emperor/empress' also lead to a disagreement in judgments among consultants:

(i) *el   duque de York y   la <duquesa> de Gran Bretaña (adapted from Saab 2010a)
    the.M duke of York and the.F duchess of Great Britain
    Intended: 'the Duke of York and the duchess of Great Britain'

For example, one other consultant and myself judge there to be a contrast with the noun pair *duque/duquesa* when the masculine is in the antecedent (v), versus the feminine (vii). For two other consultants, (v) is just as unacceptable as (vii):

(ii)   El   duque de York y el   duque de Edinburgo llegaron a Francia.
    the.M duke of York and the.M duke of Edinburgh arrive   in France
    'The duke of York and the duke of Edinburgh arrived in France.'
(iii) El duque de York y el <duque> de Edinburgo llegaron a Francia.
(iv)   El   duque de York y la   duque-sa de Edinburgo llegaron a Francia.
    the.M duke of York and the duke-SA of Edinburgh arrived in France

240

(38)   *Class III nouns*
    a. actor/actriz   'actor/actress'
    b. alcalde/alcaldesa 'male mayor/female mayor'
    c. héroe/heroína   'male hero/female hero'
    d. poeta/poetisa   'male poet/female poet'

One aspect of all nouns in this class that will become relevant concerns their morphological breakdown. In Class III nouns, the feminine version contains a morpheme that does not occur on Class I or Class II nouns. For example, the feminine noun *alcaldesa* '(female) mayor' bears the suffix /-sa/, in contrast to its masculine counterpart.

Let us observe how nouns in this class behave with respect to ellipsis and [GENDER] mismatches. Starting with NP ellipsis, note that it is possible for the masculine version to be in the antecedent:

(39)   *Class III and NPE: masculine-feminine*
    a. El   actor de Hollywood ganó un Óscar y   el   actor de
       the.M actor of Hollywood won an Oscar and the.M actor of
       Bollywood también.
       Bollywood also
       'The Hollywood actor won an Oscar and the Bollywood actor did too.'

    b. El actor de Hollywood ganó un Óscar y el <actor> de Bollywood también.
       'The Hollywood actor won an Oscar and the Bollywood one did too.'

    c. El   actor de Hollywood ganó un Óscar y   la   actr-iz de
       the.M actor of Hollywood won an Oscar and the.F actor-IZ of
       Bollywood también.
       Bollywood also
       'The Hollywood actor won an Oscar and the Bollywood actress did too.'

    d. ?El actor de Hollywood ganó un Oscar y la <actriz> de Bollywood también.
       'The Hollywood actor won an Oscar and the Bollywood (actress) did too.'

'The duke of York and the duschess of Edinburgh arrived in France.'
(v)   %El duque de York y la <duquesa> de Edinburgo llegaron a Francia.   %NPE masculine-feminine
(vi)   La   duque-sa de Edinburgo y   el duque de York llegaron a Francia.
        the.F duke-SA of Edinburgh and the duke of York arrived in France
        'The duchess of Edinburgh and the duke of York arrived in France.'
(vii)   * La duquesa de Edinburgo y el <duque> de York llegaron a Francia.   *NPE feminine-masculine

See section 4.7 for more comments on nobility terms.

241

For now, I show the ellipsis site containing *actriz* in (39)d, but I will return to an analysis of the actual content of the ellipsis site for examples of this type in section 4.4.2. The question mark in (39)d signals that the sentence is not perfect, but it is acceptable, especially in comparison to the judgment in (40)d below.

Consider, then, examples where the feminine noun is in the antecedent and its masculine counterpart is in the ellipsis site. This configuration is impossible:

(40)   *Class III and NPE: feminine-masculine*
    a. La   actr-iz de Bollywood ganó un Óscar y   la   actr-iz de
       the.F actor-IZ of Bollywood won an Oscar and the.F actor-IZ of
       Hollywood también.
       Hollywood also
       'The Bollywood actress won an Oscar and the Hollywood actress did too.'
    b. La actriz de Bollywood ganó un Óscar y la \<actriz\> de Hollywood también
       'The Bollywood actress won an Oscar and the Hollywood one did too.'
    c. La   actr-iz de Bollywood ganó un Óscar y el   actor de
       the.F actor-IZ of Bollywood won   an Oscar and the.M actor of
       Hollywood también.
       Hollywood also

       'The Bollywood actress won an Oscar and the Hollywood actor did too.'
    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 see, then, that Class III nouns behave asymmetrically when it comes to [GENDER] mismatches in ellipsis, in a way that is distinct from the behavior of such mismatches with nouns in either Class I or Class II. Note, as well, that the data just described are parallel in behavior to the Greek (16) and Brazilian Portuguese (13).

The data below bring the point home using predicate ellipsis— the contrast is identical. Again, the example in (41)d is degraded, but better than (42)d.

(41)   *Class III and predicate ellipsis: masculine-feminine*

    a. Jaime no es actor , pero Marlon sí   es actor.
       Jaime not is actor   but   Marlon yes is actor
       'Jaime is not an actor, but Marlon is an actor.'

    b. Jaime no es actor , pero Marlon sí <es actor>.
       'Jaime is not an actor, but Marlon is.'

    c. Jaime no es actor, pero Elizabeth sí   es actr-iz.
       Jaime not is actor but   Elizabeth yes is actor-IZ
       'Jaime is not an actor, but Elizabeth is an actress.'

    d. ?Jaime no es actor, pero Elizabeth sí <es actriz>.
       'Jaime is not an actor, but Elizabeth is (an actress).'

(42)   *Class III and predicate ellipsis: feminine-masculine*

    a. Laura no es actri-z, pero Elizabeth sí   es actr-iz.
       Laura not is actor-IZ but   Elizabeth yes is actor-IZ
       'Laura is not a an actress, but Elizabeth is an actress.'

    b. Laura no es actriz, pero Elizabeth sí <es actriz>.
       'Laura is not an actress, but Elizabeth is.'

    c. Laura no es actr-iz, pero Marlon sí   es actor.
       Laura not is actor-IZ 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>.
       Intended: 'Laura is not an actress, but Marlon is (an actor).'

These examples are parallel to the Brazilian Portuguese examples in (10) and the Russian in (19).

To summarize, we observe an asymmetrical pattern with Class III nouns: whereas a masculine noun in the antecedent can mismatch with its feminine counterpart in the ellipsis site, the opposite configuration is impossible. The empirical picture established in this section is summarized once again below:

(43)   *[GENDER] mismatches under ellipsis—three classes of nouns (repeated from (23))*
    a. *Class I*: [GENDER] mismatch symmetrically ill-formed   e.g. *tío ~ tía*
    b. *Class II*: [GENDER] mismatch symmetrically well-formed
                                 e.g. *abogado ~ abogada*
    c. *Class III:* [GENDER] mismatch asymmetrically well-formed   e.g. *actor ~ actriz*
        (i)   Masculine antecedent ✓
        (ii)   Feminine antecedent *

At this juncture, it is important to emphasize that the empirical picture in (43) should be particularly puzzling for any approach to the syntactic identity condition on ellipsis that would require *strict* identity. If [GENDER] is mismatching in all the data discussed here, why should three classes of nouns emerge, each exhibiting a different pattern? I will show in what follows that the approach defended in this dissertation can make sense of this three-way asymmetry.

To summarize, we established that a (by now familiar) pattern emerges in Spanish as well—there are (at least) three classes of nouns that behave distinctly under ellipsis when [GENDER] is forced to mismatch. While I will add to the empirical picture in section 4.5, we are ready to present an analysis that will capture the data presented in this section.

4.3 Setting the stage for the analysis

This section lays out the moving parts that will underlie my analysis of [GENDER] mismatches involving Class I, Class II, and Class III nouns in 4.4. The analysis pursued therein will be deeply indebted to the analyses in Saab 2004 and Saab 2008: chapter 5, whose account for a subset of the classes of nouns I will discuss is very similar (see the analysis of Class I; see also comments on Class IV, V, and VIII in 4.5)—modulo a difference in the strictness of the identity condition that is assumed (see 4.6).

Before delving into the ellipsis data and analysis in 4.4, I first go over my assumptions about NP ellipsis (4.3.1), the structure of nominals (4.3.2), repair-by-ellipsis (4.3.3), and the operation underlying nominal concord (4.3.4).

4.3.1 The derivation of NP ellipsis

The examples discussed in this chapter involve predicate ellipsis with a remnant *también* 'too' or a polarity particle *sí/no* 'yes/no' (see chapter 2 for an analysis of the latter based

on Saab 2010b) and NP ellipsis. Focusing now on the latter, I will follow the analysis proposed in Saab 2010a, 2019 for this elliptical construction: NP ellipsis involves ellipsis of the *n*P node. What this entails is that the Num^0 head bears the [E] feature and licenses ellipsis of its complement. We will delve more closely into the featural breakdown of the DP in 4.3.2 below:

(44)   *NP ellipsis* (based on Saab 2010a)

```
             DP
           /    \
          D     NumP
               /    \
         Num[E] - - - nP         ← NP ellipsis targets nP
                    /    \
                   n      √
```
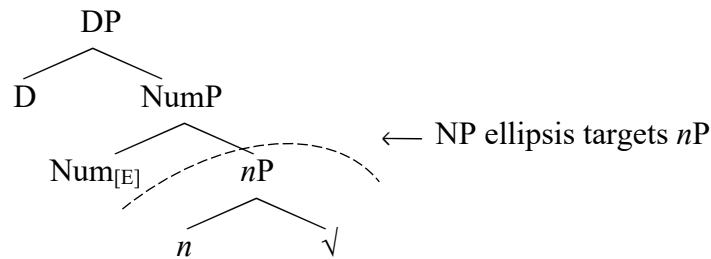
Note that the analysis in (44) straightforwardly explains why [NUMBER] is allowed to mismatch freely under NP ellipsis: in this construction, the Num^0 head is outside of the ellipsis site, so the [NUMBER] feature is not taken into account when calculating the satisfaction of the identity condition on ellipsis (see a similar logic for the analysis of well-formed voice mismatches under VP ellipsis in chapter 2 from Merchant 2013).

At this time, it is necessary to establish that the gap in the examples that I have labelled NP ellipsis is truly derived via *ellipsis*; in other words, we should ensure that the gap is a *surface anaphor* in the sense of Hankamer & Sag 1976, as opposed to being a null pronominal, a *deep anaphor* (see chapter 1). Perhaps the most compelling piece of evidence for the analysis of NP ellipsis as true ellipsis involves the possibility of sub-extraction from the gap[103]. This diagnostic has a long tradition in the literature (e.g., Fiengo & May 1994)

---

[103] See Saab 2019 for other evidence like (i) matching effects and (ii) non-ambiguous readings under NP ellipsis versus ambiguous readings for null pronominals.

and has proven useful in establishing that there is underlying structure in other elliptical constructions such as VP ellipsis, which allows sub-extraction, whereas a surface similar construction known as null complement anaphora (NCA) does not allow sub-extraction (Depiante 2000, Depiante 2019; see chapter 3 for an application of this diagnostic to Spanish predicate ellipsis). The logic of the diagnostic is the following: if extraction is possible from a phonetically null element, then there is structure inside that element (recall that I take ellipsis to involve PF-deletion (or lack of insertion); see chapter 1):[104]

(45)   *The sub-extraction diagnostic: VPE vs. NCA (adapted from Merchant 2013b)*
         a. Which films did he refuse to see and [which films]$_1$ did he agree to <see t$_1$> ?
                                                                                              VP ellipsis
         b. *Which films did he refuse to see and which films did he agree?   *NCA*

Consider now the sub-extraction diagnostic as applied to the examples that concern us here. As shown below in (46)b, Spanish gaps of the sort I have labelled NP ellipsis allow for sub-extraction (see Saab 2019 for more examples from Spanish and Hungarian; see Merchant 2014, Sudo & Spathas 2016 for Greek):

(46)   *NP ellipsis and sub-extraction (adapted from Saab 2019)*
         a. Yo sé   de quién compraste dos fotos, pero no sé   [de quién]$_1$ compraste
            I   know of who   bought.2s two photos but   not know of who   bought.2s
            tres [fotos t$_1$].
            three photos
            'I know who you bought two photos of, but I don't know who you bought three photos of.'
         b. Yo sé de quién compraste dos fotos, pero no sé [de quién]$_1$compraste tres <[fotos t$_1$]>.

---

[104] The converse—that the impossibility of sub-extraction entails the absence of structure—is more controversial; see Aelbrecht 2010, Thoms 2019, Depiante 2019.

Crucially, the examples below show that sub-extraction is possible when we force a [GENDER] mismatch for Class II nominals; (47)c-d and (48)c-d. This is specific evidence that the examples we are assessing involve ellipsis and not a null-pronominal:[105]

(47)   *NP ellipsis, sub-extraction, and [GENDER] mismatch: Class II masculine-feminine*

    a. Yo sé   de quién es este abogado, pero no sé   de quién es
    I   know of who is this.M lawyer.M but not know of who   is
    aquella abogada.
    that.F   lawyer.F
    'I know whose (male) lawyer this is, but I don't know whose (female) lawyer that is.'

    b. Yo sé [de quién]$_1$ es este abogado t$_1$, pero no sé [de quién]$_2$ es aquella abogada t$_2$.

    c. Yo sé de quién es este abogado, pero no sé de quién es aquella.
    'I know whose (male) lawyer this is, but I don't know whose (female lawyer) that is.'   *NP ellipsis*

    d. Yo sé [de quién]$_1$ es este abogado t$_1$, pero no sé [de quién]$_2$ es aquella <abogada t$_2$>.

(48)   *NP ellipsis, sub-extraction, and [GENDER] mismatch: Class II feminine-masculine*

    a. Yo sé   de quién es esta abogada, pero no sé   de quién es
    I   know of who is this.F lawyer.F but not know of who   is
    aquel abogado.
    that.M lawyer.M
    'I know whose (female) lawyer this is, but I don't know whose (male) lawyer that is.'

    b. Yo sé [de quién]$_1$ es esta abogada t$_1$, pero no sé [de quién]$_2$ es aquel abogado t$_2$.

    c. Yo sé de quién es esta abogada, pero no sé de quién es aquel.
    'I know whose (female) lawyer this is, but I don't know whose (male lawyer) that is.'   *NP ellipsis*

    d. Yo sé [de quién]$_1$ es esta abogada t$_1$, pero no sé [de quién]$_2$ es aquel <abogado t$_2$>.

---

[105] Even if the evidence led us to conclude (contrary to fact) that the examples I labelled NP ellipsis are deep anaphora, we would still be faced with a puzzle: the pattern of [GENDER] mismatches in deep anaphora would be contingent on membership into one class of nouns or another. The solution to that hypothetical puzzle, however, would have no bearing on the identity condition *on ellipsis*; rather, it would have a bearing on some other area of the grammar.

As expected, sub-extraction is also possible in well-formed [GENDER] mismatches with Class III nominals. Note that I have filled in the ellipsis site with *actriz* in (49)c below, but my analysis of the content of the ellipsis site here will be different; see 4.4.2:

(49)   *NP ellipsis, sub-extraction, and [GENDER] mismatch: Class III masculine-feminine*

Context: we are at a movie set where two movies are being filmed by different directors and there are actors and actresses around. I know who one actor is working for, but I don't know who one actress is working for.

a. Yo sé   de quién es este   actor, pero no sé   de quién es
   I   know of who   is this.M actor but   not know of who   is
   aquella actriz.
   that.F   actress
   'I know whose actor this is, but I don't know whose actress that is.'

b. Yo sé [de quién]$_1$ es este actor t$_1$, pero no sé [de quién]$_2$ es aquella actriz t$_2$.

c. ?Yo sé de quién es este actor, pero no sé de quién es aquella.
   'I know whose actor this is, but I don't know whose (actress) that is.'
                                                                    NP ellipsis

d. Yo sé [de quién]$_1$ es este actor t$_1$, pero no sé [de quién]$_2$ es aquella <actriz t$_2$>.

Observe, for completeness, that a [GENDER] mismatch with sub-extraction is ill-formed when the feminine is in the antecedent:[106]

(50)   *NP ellipsis, sub-extraction, and [GENDER] mismatch: Class III masculine-feminine*

a. Yo sé   de quién es esta   actriz, , pero no sé   de quién es
   I   know of who   is this.F actress but   not know of who   is
   aquel   actor.
   that.M actor
   'I know whose actress this is, but I don't know whose actor that is.'

b. Yo sé [de quién]$_1$ es esta actriz t$_1$, pero no sé [de quién]$_2$ es aquel actor t$_2$.

c. *Yo sé de quién es esta actriz, pero no sé de quién es aquel.
   Intended: 'I know whose actress this is, but I don't know whose (actor) that is.'

d. *Yo sé [de quién]$_1$ es esta actriz t$_1$, pero no sé [de quién]$_2$ es aquel <actor t$_2$>.

---

[106] One also cannot mismatch [GENDER] with Class I nouns and sub-extract from the gap. An example like the following is impossible:

(i)   *Yo sé   de quién es este   gato, pero no sé   [de quién]$_1$ es aquella <gata t$_1$>.
          I   know of who   is this.M cat.M but   not know of who   is that.F   cat.F
          Intended: 'I know whose (male) cat this is, but I don't know whose (female cat) that is.'

To summarize this subsection, I have presented evidence that NP ellipsis involves true ellipsis; i.e., there is structure in the ellipsis site that is subject to the identity condition. With this established, we can move on to the next set of assumptions which underlie the analysis presented in 4.4.

4.3.2 The structure of nominals

In this subsection, I lay out my assumptions regarding the locus of [GENDER] in Spanish. The literature on the decomposition of nominals and the specific location of [GENDER] within the nominal spine has been a fertile area of research in recent years, in particular within the Distributed Morphology framework (see Kramer 2015, Kučerová 2018, Mathieu et. al. 2019, Kramer 2020, and references therein). However, it is not my intention here to use ellipsis to adjudicate between all of the proposals currently on the market for the locus of [GENDER]. Rather, my goal here is to show how a specific proposal about the locus of [GENDER]— with independent support in the literature—can form one of the pillars of an analysis of [GENDER] mismatches under ellipsis. The other pillars will be the syntactic identity condition proposed in this dissertation and certain ancillary assumptions about the repair capability of ellipsis (see 4.3.3).

Concretely, I assume a decompositional view of nominals couched in Distributed Morphology (DM; Halle & Marantz 1993, Arad 2005, Embick & Noyer 2007, Harley 2014, a.o.). In particular, (underived) nominals are composed of several layers of structure where the innermost element is an acategorial $\sqrt{\text{ROOT}}$ that is categorized by a nominalizing head n^0. In early versions of DM, $\sqrt{\text{ROOT}}$s were not considered to be individuated syntactically (Marantz 1995; see Harley 2014 for discussion). I depart from this view and assume that $\sqrt{\text{ROOT}}$s are featureless syntactic primitives that are individuated in the syntax

via an index that instructs the interpretive (Encyclopedia) and phonological (Vocabulary Insertion) components post-syntactically (following Harley 2014, Kramer 2015; for an alternative proposal that takes √ROOTs to be phonologically individuated, see Borer 2005, 2014). Consider as an illustration the noun 'pepper' (51):

(51)   *√ROOT categorization*

$$nP$$

         n   √47
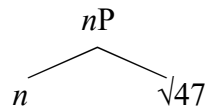
In this structure, the root √ is individuated in the syntax via an index (here, 47) and instructs the post-syntactic components for insertion (in a particular structural context; e.g. in the context of a nominalizing head) of phonological and semantic content:[107]

(52)   *√47 post-syntax; nominal context*
         Vocabulary Insertion: √47 ↔ [ˈpɛ.pəɹ]
         Encyclopedia: $[_{nP}$ [n [√47]] is interpreted as a spice, used to season food, which can be of different kinds (black, pink)…

I will come back to √ROOTs and a specific proposal that some affixes are √ROOTs (Lowenstamm 2015, Creemers et al. 2018) in 4.4.2, where I discuss Class III mismatches.

   Having established my basic assumption about √ROOTs, we can move on to the locus of [GENDER] and [NUMBER] in the nominal spine. I assume that [GENDER] is located on the nominalizing head n^0 (Kramer 2015). Moving one level higher in the structure, I assume that [NUMBER] features are introduced on Num^0 (Ritter 1991; see also Preminger 2020 and

---

[107] Note that I am using Kramer's notation for the Encyclopedia. An alternative way to notate the insertion of √ROOTs in the semantic component would be as follows (see, for example, (69)):

(i)   *Encyclopedia*:   √47 ↔ "a spice, used to season food…" / $[_{nP}$ [n _]]

references therein), which is immediately above n^0. Finally, I assume that nominals are headed by D^0 (Abney 1987 and much subsequent work):

(53)   *Nominal structure*

```
              DP
            /    \
          D       NumP
                 /    \
              Num      nP
                      /   \
                     n     √
                 [φ: GENDER]
```
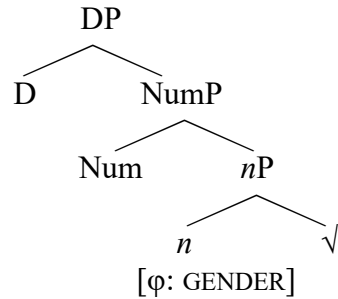
I also assume that n^0 comes in different flavors—i.e., it can be specified for different [GENDER] features. This will be crucial for the analysis hereafter. For our current purposes, let us assume that there are three distinct flavors of n^0 in Spanish: one bearing interpretable i[+F(EMININE)], another bearing interpretable i[-F], and a third bearing no [GENDER] feature at all, which I will call default or plain n:

(54)   *[GENDER] in Spanish (to be revised; partially from Kramer 2015)*
    a. The locus of [GENDER] is the categorizing head n^0.
    b. Default grammatical gender is masculine.
    c. There are three flavors of n:
       (i) $n_{i[+F]}$ (feminine)
       (ii) $n_{i[-F]}$ (masculine)
       (iii) n (default; masculine concord is triggered)

I assume that in the semantic component, i[+F] and i[-F] contribute to the denotation. In the case of human denoting nouns, i[+F] contributes the interpretation that the individual(s) uses feminine pronouns; in the case of animals, that the being is female. Moving on to i[-F], this feature contributes the interpretation that the individual(s) uses masculine pronouns; in the case of animals, that the being is male (and see below for [GENDER] on inanimate nouns).

My assumption that the default grammatical gender in Spanish is masculine has a long tradition in the literature (Prado 1982, Roca 1989; see Kramer 2015 and references therein). A question that might arise at this juncture is why a three-way opposition is needed for [GENDER], instead of assuming that there is a female $n_{i[F]}$ and masculine/default is a plain n which bears no [GENDER] feature. The demonstrative system in Spanish illustrates why such a three-way distinction is needed:

(55)   *Spanish demonstrative triplet*
    a. aquel garrobo   'that (male) iguana'
        that.M male.iguana
    b. aquella iguana   'that (female) iguana'
        that.F   female.iguana
    c. aquello$_{pro}$ 'that (thing/situation)'
        that

Demonstrative triplets like the above show why a three-way distinction is required, rather than a simple opposition between a feature and its absence; another distal demonstrative triplet *ese*, *esa*, *eso* and the proximate demonstrative triplet *este*, *esta*, *esto* illustrate the same point. A putative alternative system where only a two-way featural distinction was proposed between $n_{i[F]}$ and plain n could not account for these three different types of demonstratives. Moving forward, then, I will assume that there is a distinction between i[+F], i[-F], and the absence of a [GENDER] feature, even though there might occasionally exist some degree of syncretism between masculine as i[-F] and masculine as the absence of [GENDER], a fact which can obscure the three-way contrast.[108]

With this in mind, we can now enrich our discussion of [GENDER] and turn to *inanimate* nouns. The grammatical [GENDER] feature borne by Spanish inanimate nouns is completely

---

[108] Thank you to Omer Preminger for discussing with me the relevance of these facts.

arbitrary and does not correlate with any interpretation. Consequently, I will follow Kramer 2015, 2020's proposal that inanimate, grammatically feminine nouns, are licensed by a nominalizer bearing a [GENDER] feature that is uninterpretable—namely, $n_{u[+F]}$. Meanwhile, inanimate, grammatically masculine nouns are licensed by default n. I update our catalogue of n^0 heads in Spanish to reflect my adherence to this proposal:[109]

(56)   *[GENDER] in Spanish (Kramer 2015)*
  a. The locus of [GENDER] is the categorizing head n^0.
  b. Default grammatical gender is masculine.
  c. There are four flavors of *n*:
   (i) $n_{i[+F]}$ (feminine)
   (ii) $n_{i[-F]}$ (masculine)
   (iii) *n* (default; controls masculine concord)
   (iv) $n_{u[+F]}$ (uninterpretable feminine)

For example, in the alternation *el caso* 'the case' and *la casa* 'the house', the √ROOT in the latter is categorized by a n^0 bearing an uninterpretable feminine feature u[+F]. This feature does not contribute anything related to human gender or animal sex to the denotation of the nominal. Nevertheless, this feature controls feminine concord on determiners and modifiers. I will come back to this fourth *n* head when I discuss inanimate nouns (Class VIII) and their behavior under ellipsis in section 4.5.4.

Having established my assumptions about [GENDER] in Spanish, we can now turn to the interaction of √ROOTs and n^0. I follow Kramer 2015 in assuming that √ROOTs are *licensed* by particular flavors of n^0. By licensing, I mean that there exist restrictions on the flavor(s) of n^0 that can combine with specific √ROOTs. Take, for example, the noun *nuera* 'daughter-in-law'. I follow Kramer 2015 in assuming that this kind of noun is composed

---

[109] I adopt Kramer's analysis, but there exist proposals that do away with uninterpretable [GENDER] features altogether (e.g., Hammerly 2019); see section 4.5.4 for a possible alternative.

of a √ROOT —say √23— that can only combine with $n_{i[+F]}$. Combining this √ROOT with any other flavor of nominalizing head is a violation of the licensing conditions specific to the √ROOT, signaled below by an asterisk:

(57)   *Licensing conditions for √23—nuera 'daughter-in-law'*
   a.   nP
         ⟨ ⟩
      $n_{i[+F]}$ √23
   b. *   nP
         ⟨ ⟩
      $n_{i[-F]}$ √23
   c. *   nP
         ⟨ ⟩
        n   √23
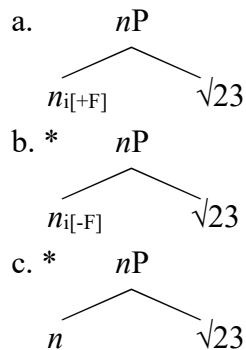
The question that arises here is what the source of the violations in (57)b and (57)c is. In other words, where in the grammar do these configurations create a problem?

   To set the stage for an answer to this question, consider how the syntax of (57)a—the only possible combination of √23 with a nominalizing head—is read at the interfaces:

(58)   *√23 post-syntax—nuera 'daughter-in-law'*
   Vocabulary Insertion: √23 ↔ /ˈnwe.ɾa/
   Encyclopedia: [$_{nP}$ [$n_{i[+F]}$ [√23]] is interpreted as an individual, who uses feminine pronouns…

I endorse the proposal that licensing violations can be the result of a problem in the Encyclopedic module of the grammar (see Kramer 2015: 51-54 and below on violations that are not semantic in nature). As I will show, this approach will shed light on mismatches that satisfy the identity condition, but are nevertheless ill-formed.

   What this proposal boils down to is that no interpretation is possible for certain combinations of nominalizing heads and a √ROOT. In the case of our noun *nuera* 'daughter-

254

in-law', this amounts to the following—if √23 combines with $n_{i[-F]}$ or plain n in the syntax, an issue arises post-syntactically; no interpretation is possible for the structure:

(59)   *√23 post-syntax—nuera 'daughter-in-law'*
    a. *Encyclopedia*: $[_{nP} [n_{[-F]} [√23]]$ is interpreted as ???   (57)b
    b. *Encyclopedia*: $[_{nP} [n$ $[√23]]$ is interpreted as ???   (57)c

Throughout the rest of this chapter, I will use the term licensing violation (notated in structures like in (57) via an asterisk *) to mean that the structural configuration under assessment would lead to a problem in the interpretive component of the grammar.

I am departing from Kramer 2015's proposal on a notational, but not substantive level. Specifically, Kramer proposes that licensing violations can arise for *two* reasons. The first is the reason just outlined: a combination of a specific nominalizer and a specific √ROOT leads to a semantic problem. The second reason, Kramer proposes, is that a combination of a specific nominalizer and a specific √ROOT would lead to an issue in the *Vocabulary Insertion* component of the grammar—what Kramer calls *arbitrary* licensing conditions.[110] In other words, Kramer proposes that it is possible that a violation of a licensing condition creates a problem for the externalization component of the grammar. Consider as an illustration a hypothetical √ROOT that can only combine with $n_{i[+F]}$:

(60)   *Hypothetical √33*
    Vocabulary Insertion: $√33 ↔ /ˈba.ba.za/ / n_{i[+F]} \_$
    Encyclopedia: $[_{nP} [n_{i[+F]} [√33]]$ is a female animal that runs
                                 around…
Kramer's second conception of a licensing violation would amount to the following for our hypothetical √ROOT: a combination of √33 with $n_{i[-F]}$ or n does not create a *semantic*

---

[110] Kramer notes that this particular conception of licensing violation is usually the one that is assumed, not the conception wherein a semantic problem occurs (see Harley & Noyer 1999, Siddiqi 2009).

problem. An issue arises, however, in the externalization component—i.e., there is a morphophonological gap and no Vocabulary Item is inserted:

(61)   *Hypothetical √33*
   a.
   Vocabulary Insertion: √33 ↔ ??? / $n_{i[\text{-F}]}$ _
   Encyclopedia: [$_{n}$P [$n_{i[\text{-F}]}$ [√33]] is a male animal that runs around…
   b.
   Vocabulary Insertion: √33 ↔ ???
   Encyclopedia: [$_{n}$P [n [√33]] is an animal that runs around…

For ease of exposition, I will not refer to cases like the above as licensing violations. Rather, I will refer to them as morphophonological gaps.

The distinction between what I will exclusively call licensing violations—a combination of nominalizer and √ROOT resulting in a post-syntactic problem in the Encyclopedia (59)—versus morphophonological gaps—combinations of nominalizer and √ROOT resulting in an externalization problem (61)—will be a key component in my analysis. As I will discuss in 4.3.3, independent evidence in the literature has suggested that ellipsis can frequently repair externalization problems—e.g., morphophonological gaps—given that ellipsis is the literal lack of externalization.[111] In contrast, insertion violations whose source does not lie in the externalization component cannot be repaired by ellipsis. This distinction will be fundamental for the derivation of possible and impossible [GENDER] mismatches for Class III in the following section, and will play a key role in section 4.5, where I expand the empirical picture:

---

[111] Some nuance, of course, is needed here: some ellipsis types (e.g., contrastive fragments answers) do not ameliorate island effects.

(62)   *Repair-by-ellipsis: assumptions moving forward*
        a. Morphophonological gaps can be repaired by ellipsis
        b. Encyclopedic gaps cannot be repaired by ellipsis

In a nutshell, if the analysis in this chapter is on the right track, the range of possibilities we observe regarding [GENDER] mismatches under ellipsis provides support for the notion that certain gaps can exist under ellipsis, while others cannot (Mendes 2020, Mendes & Nevins 2021, Mendes & Kandybowicz 2021 and references therein).

To summarize this subsection, I laid out my assumptions about the structure of nominals, the nature of √ROOTs, and the location of [NUMBER] and [GENDER] in the nominal spine. I adopted Kramer 2015's proposal in assuming that [GENDER] is on the nominalizing n^0, which in Spanish comes in four flavors. I also described the nature of licensing conditions between n^0 and specific √ROOTs, making a distinction between licensing violations—which I take to be semantic in nature—and morphophonological gaps. With that in mind, we now take a brief detour into repair-by-ellipsis (4.3.3) and nominal concord (4.3.4), before pivoting back to the ellipsis data in 4.4.

4.3.3 Repair-by-ellipsis

In this subsection, I discuss repair-by-ellipsis (also called salvation-by-deletion), delving into the distinction between grammatical violations that are allowed under ellipsis, and those that are not allowed under ellipsis. The possibility or impossibility of repair-by-ellipsis will play a role in the derivation of well-formed masculine-feminine mismatches with Class III. It will also explain why certain derivations that satisfy featural non-distinctness are nevertheless ill-formed (e.g., Class I feminine-masculine mismatches in 4.4.1 where the ellipsis site contains plain *n*).

There exists a rich tradition of work on the repair power of ellipsis, dating back to the seminal work of Ross 1969a, with a particular focus being placed on the issue of island repair via sluicing (Lasnik 2009; see chapter 2). In short, I will follow Mendes 2020, Mendes & Nevins 2021, and Mendes & Kandybowicz 2021 and assume that deviances that have a morphophonological underpinning are repaired by ellipsis, whereas deviances that arise due to issues unrelated to externalization cannot be repaired by ellipsis. Put differently, if the source of a deviance or lexical gap is due to a post-syntactic issue in the externalization component, then ellipsis can repair the deviance. This follows naturally, since ellipsis is definitionally the absence of externalization. In contrast, if a structure is (i) not derivationally available or (ii) would lead to a problem in the semantic component of the grammar, then ellipsis has no repair power, since the issue does not lie in externalization proper.

To illustrate the repair power of ellipsis, let us delve into the phenomenon wherein certain lexical gaps are extraordinarily allowed under ellipsis (see Kennedy & Merchant 2000, Abels 2019b). For instance, consider the present tense paradigm for the Spanish verb *abolir* 'to abolish' (see Mendes & Nevins 2021 for similar discussion of the Brazilian Portuguese verb *demolir* 'to demolish'). As shown below, there are several gaps in this paradigm—speakers simply cannot produce any form for a subset of the cells:

(63)   *Morphological gaps—Spanish present tense 'abolir'*

1s	yo	-	*Intended:* 'I abolish'
2s	vos	abolís	'you abolish' (informal)
2s	tú	-	*Intended*: 'you abolish'
2s	usted	-	*Intended*: 'you abolish' (formal)
3s	él/ella	-	*Intended:* 'he/she abolishes
1P	nosotros	abolimos	'we abolish'
2P	ustedes	-	*Intended*: 'you all abolish'
3P	ellos/ellas	-	*Intended*: 'they abolish'

Taking the 1S cell as representative of the gaps throughout the paradigm, none of the conceivable forms for the intended exponent are judged as acceptable:

(64)   *Morphological gaps—Spanish present tense 'abolir'*
     *Yo abolo/abuelo las leyes.
      I   abolish.1S   the laws
     Intended:* 'I abolish the laws.'

However, it *is* possible to express the thought intended in (64) inside a gap derived via ellipsis. Consider the examples below, which illustrate this repair effect of ellipsis:

(65)   *Morphological gap repair under ellipsis*
     a. *Vos abolís   las leyes por   malicia, pero yo no **abuelo/abolo**
       you abolish.2S.PRS the laws because malice but   I   not abolish.1S.PRS
       las leyes por   malicia.
       the laws because malice
       Intended: 'You abolish the laws out of malice, but I do not abolish the laws out
          of malice.'
     b. Vos abolís las leyes por malicia, pero yo no <**abuelo/abolo** las leyes por
       malicia>.
       'You abolish the laws out of malice, but I don't.'

I follow Mendes & Nevins 2021 in analyzing repair effects like the above as follows: in the post-syntactic Vocabulary Insertion component (see chapter 2), certain combinations of (i) present tense features, (ii) person/number features, and (iii) the relevant √ROOT yield no output—i.e., the gaps in the paradigm in (63) and the deviance of the verb forms in (65)a arise due to a morphophonological issue. However, since ellipsis is the literal absence of externalization, we can explain the well-formed status of the *elliptical* (65)b.

In contrast to the phenomenon just described, some deviances do not seem to be repairable by ellipsis (see Mendes 2020, Mendes & Nevins 2021, Mendes & Kandybowicz 2021 for discussion). One example of deviances that seem to resist repair under ellipsis was already discussed in chapter 2: preposition stranding violations. Other cases that have been discussed in the literature include Empty Category Principle (ECP) violations (Nakao

2009), Head Movement Constraint (HMC) violations (Mendes 2020), and Superiority violations (Boeckx and Lasnik 2006). Taking the last case as an illustration, Boeckx & Lasnik (2006) discuss how Serbo-Croatian sluicing does not seem to repair this type of violation, which they contrast with *wh*-islands, which do seem to be repairable (see chapter 2 here for a discussion of island repair via sluicing). Consider the baseline data below in (66):

(66)   *Superiority; Serbo-Croatian (adapted from Boeckx & Lasnik 2006)*
        Ivan i   Marko ne   znaju…
        Ivan and Marco NEG know
                a. ko   je šta   kupio.   who > what
                   who is what bought
                b. *šta   je ko   kupio.   *what > who
                    what is who bought
                'Ivan and Marko don't know who bought what.'

As shown below, sluicing does not repair the Superiority violation:

(67)   *Superiority; Serbo-Croatian sluicing (adapted from Boeckx & Lasnik 2006)*
        a. Somebody bought something, but…
        b.   i.   Ivan i   Marko ne   znaju ko   šta.   who > what
                     Ivan and Marko NEG know who what
                ii. *Ivan i   Marko ne   znaju šta   ko.   *what > who
                     Ivan and Marko NEG know what who
                     'Ivan and Marko don't know who what.'

The data presented by Boeckx and Lasnik seem to show that ellipsis—an interface condition that is definitionally the absence of externalization—cannot fix a violation that has nothing to do with the externalization component; i.e., Superiority violations are *syntactic* and thus still incurred under ellipsis.[112]

---

[112] In contrast to the picture in Serbo-Croatian, see Grebenyova 2006, who shows that Russian Superiority facts under sluicing are quite different.

Another phenomenon whose analysis will be relevant for our discussion involves lexical gaps in the context of idiomatic *pluralia tantum* nouns. As shown by Mendes & Nevins 2021, an expression like *high jinks* (i.e., mischief)—which is exclusively grammatically plural (*high jink*)—cannot participate in a [NUMBER] mismatch under ellipsis. Observe the example below, which shows that attempting to elide a singular version of the expression *high jinks* is ill-formed. Recall, crucially, that [NUMBER] mismatches are allowed under ellipsis (data adapted from Mendes & Nevins 2021):

(68)   *Idiomatic pluralia tantum and [NUMBER] mismatch*
         a. *I don't care for these high jinks, not even one <high jink>.
         b. *I don't care for John's high jinks, especially the last <high jink>.

The ill-formed status of (68)a,b contrasts with the well-formed status of examples like in (65). I will adopt Mendes & Nevins 2021's account for this contrast: in a nutshell, the source of the gaps in the present tense paradigm of the verb *abolir* 'to abolish' is not the same as the source of the gap for singular *high jinks*. Mendes & Nevins follow Harley 2014 in proposing that the issue with a grammatically singular version of the expression *high jinks* (elided in (68)b above) does not lie in the morphophonological component—in contrast to the source for the gaps in (64). Rather, there exists a problem on the semantic side of the grammar—there is no encyclopedic entry for a structure in which the root √JINK does not occur in the environment of a [+plural] feature and the adjective *high* (i.e., *low jinks* is not well-formed):

(69)   *Encyclopedic entry for jinks (Mendes & Nevins 2021)*
         √JINK ↔ mischief′ / [DP high [#P [nP [_ n] [+plural]]]   (*no elsewhere item*)

Since the issue with the elided singular counterpart of *high jinks* has nothing to do with externalization, then the lack of repair by ellipsis in (68) is expected.

To summarize, I discussed the phenomenon of repair-by-ellipsis and showed evidence that morphophonological gaps are repaired by ellipsis, whereas syntactic or semantic violations cannot be. In what follows, I will use repair (and non-repair) by ellipsis as a key ingredient to derive the patterns we observed regarding [GENDER] mismatches with nouns in different classes.

4.3.4 A note on nominal concord

The identity condition defended thus far in this dissertation specifies that the antecedent and elements that are *properly contained* in the ellipsis site must be featurally non-distinct. I used the proper containment clause to explain why a *wh*-remnant in sluicing, as well as [PERSON] mismatches in predicate ellipsis, do not lead to a violation of the condition.

Given this proposal, though, a question arises at this juncture: why aren't [GENDER] mismatches in NP ellipsis well-formed across the board, since a determiner or modifier in the remnant bears a morphological reflex of the [GENDER] feature borne by the nominal in the ellipsis site?

(70)   *Class I and NPE: feminine-masculine (see (28))*
     *La   gata de María es dócil, pero **el**   <**gato**> de Juan es feroz.
     the.F cat.F of María is docile but **the.M. cat.M** of Juan is ferocious
     Intended: 'María's (female) cat is docile, but Juan's (male) cat is ferocious.'

What interests us in this sub-section is determining whether the mechanism underlying nominal concord in the DP involves feature sharing or not—i.e., whether the masculine [GENDER] feature in (70) is shared between the concord controller *gato* '(male) cat' and the determiner. In other words, we need to assess whether nominal concord is the result of Agree (Chomsky 2000, 2001; see D'Alessandro 2021), which I assumed involves feature sharing (see chapter 2), or not. Indeed, there are many proposals that *do* seek to analyze

nominal concord as the result of Agree within the DP, with slight variations (e.g., Schoorlemmer 2009, Toosarvandani & van Urk 2013, Carstens 2016, Landau 2016, a.o.; see Norris 2017 for a discussion of differences between these kinds of approaches).

However, empirical generalizations have accumulated in the literature that suggest that subject-predicate agreement and nominal concord should be treated as resulting from distinct operations. Norris 2014, 2017 summarizes four asymmetries that cannot be straightforwardly explained if subject-verb agreement and nominal concord are both underpinned by Agree. First, there is a difference in the number of loci where the agreeing feature is expressed: in subject-verb agreement, the feature is usually expressed once, whereas in concord, it can be expressed an unbounded number of times within the same DP. Second, the structural position of elements that expone concord is not uniform—heads can bear concord (e.g., determiners), specifiers can as well (e.g., possessors), as can adjuncts (adjectives). In subject-verb agreement, in contrast, heads usually expone agreement. Third, subject-verb agreement is a relation between two different extended projections—the agreement controller is a nominal projection, and agreement is expressed on a verbal projection. In contrast, the controller and bearer of concord are in the same (nominal) extended projection. Finally, subject-verb agreement can be conditioned by the morphological case of the controller—e.g., only nominals bearing nominative case control agreement in Estonian—whereas no such sensitivity to case has been established in nominal concord.

I take seriously the empirical picture just outlined and assume that the evidence is strong for the viewpoint that concord is not the result of Agree. I do not commit to a specific proposal for the operation underlying concord, but note that there are several alternatives.

For example, Polinsky (2016)b discusses how subject-verb agreement in Archi is constrained by locality—a clear indication that Agree is involved—in contrast to nominal concord, which is unconstrained by the same conditions. Polinsky proposes as a result that concord involves an operation of φ-feature *copying*. Another proposal is Norris 2014's. In analyzing the intricate pattern of concord in Estonian, Norris proposes that concord involves feature *percolation* throughout the nominal spine, with φ-features like [NUMBER] spreading upward and case-features spreading downward.

In what follows, then, I will assume that concord is a separate operation from Agree and does not involve feature sharing in the same way as Agree does. Naturally, if convincing arguments arose that concord *must* involve Agree—i.e., the empirical asymmetries raised by Norris 2017 can be derived via an Agree mediated derivation—then the ill-formed status of certain [GENDER] mismatches in NP ellipsis becomes challenging for the identity condition proposed in this dissertation. In this hypothetical scenario, it would be necessary to rethink the proper containment clause of the identity condition and its role as an explanation for a subset of the data analyzed in chapter 3.

Having finished our detour into nominal concord, we are now in a position to put all the pieces together and analyze the [GENDER] mismatch patterns that we established.

4.4 Analysis

I present my analysis in this section, putting together the ingredients just introduced alongside the identity condition that has been defended thus far. In 4.4.1, I discuss Class I and II mismatches. In 4.4.2, I delve into Class III and propose a second statement for the identity condition. The addition of this statement results in the final proposal in this dissertation:

(71) *Syntactic identity condition on ellipsis (final)*
    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.

4.4.1 Class I and Class II mismatches

Let us begin by laying out the analysis that I propose for Class I nominals. For ease of exposition, I will occasionally notate √ROOTs with their phonological content or their Class number—recall, however, that I am assuming that √ROOTs bear no features and are individuated in the syntax by an index (Harley 2014).

Consider the root √TÍ, which underlies the Class I nominal pair *tío/tía* 'uncle/aunt'. I propose that this √ROOT can be licensed under (i) $n_{i[+F]}$, yielding feminine *tía* 'aunt' or (ii) $n_{i[-F]}$, yielding masculine *tío* 'uncle' (I discuss theme vowels at the end of this section). Crucially, though, I propose that √TÍ *cannot* be licensed by default *n*, which lacks [GENDER] features altogether (recall that licensing conditions refer to the availability of an encyclopedic insertion rule):

(72) *Licensing conditions for √TÍ*
    a.   *n*P   *tía* 'aunt'

          $n_{i[+F]}$ √TÍ

    b.   *n*P   *tío* 'uncle'

          $n_{i[-F]}$ √TI

    c. *   *n*P   *Intended*: '(default) uncle/aunt'

          n   √TÍ
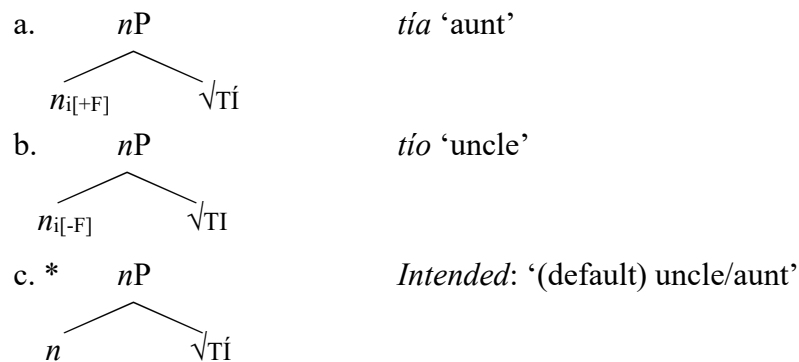
I propose that this licensing configuration holds for all Class I nominals: √ROOTs in this class are licensed under $n_{i[+F]}$ or $n_{i[-F]}$, but not under *n*:

(73)   *Licensing conditions for Class I nominals*

a. nP

$n_{i[+F]}$ √I

b. nP

$n_{i[-F]}$ √I

c. * nP

n √I
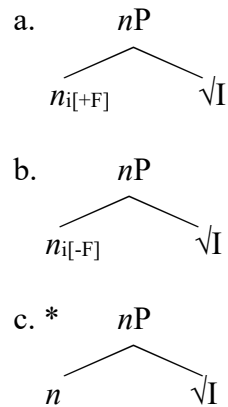
In order to derive the ellipsis patterns we observed, recall the identity condition that has

been proposed thus far in this dissertation:

(74)   *Syntactic identity condition on ellipsis (repated from (2))*
        The antecedent and material properly contained in the ellipsis site must be
        featurally non-distinct.

Mismatches allowed by the condition in (74) can be summarized abstractly as follows:

whereas clashes between features are banned under ellipsis, a mismatch between a

featurally specified node and an equivalent node lacking the relevant features is allowed.

The table below illustrates possible well- and ill-formed mismatches given the condition:

(75)   *Featural non-distinctness under ellipsis*

Antecedent	Ellipsis Site	Status
$H_{[F:\,X]}$	$H_{[F:\,Y]}$	*
$H_{[F:\,Y]}$	$H_{[F:\,X]}$	*
$H_{[F:\,X]}$	$H_{\varnothing}$	✓
$H_{\varnothing}$	$H_{[F:\,X]}$	✓

Recall that [GENDER] mismatches with Class I nouns are symmetrically ill-formed. The

proposal in (73) for the licensing conditions of Class I and the identity condition in (74)

derive the pattern, since a featural clash occurs between the antecedent and ellipsis site. I illustrate below with NP ellipsis, but the same analysis transfers directly to ellipses targeting a larger chunk of structure containing the whole DP:

(76)　*[GENDER] mismatch with Class I nouns violates featural non-distinctness*

　　a.　　　　　*Antecedent*　　　　　　　　　　　*Ellipsis site*

* nP　　　　　　　　　　　　　　　　nP

$n_{i[+F]}$　　√TÍ　　　　　　　　$n_{i[-F]}$.　　√TÍ

　　La tía de Ana y el <tío> de Clara llegaron.
　　Intended: 'Ana's aunt and Clara's (uncle) arrived.'

(repeated from (31)h)

　　b.

* nP　　　　　　　　　　　　　　　　nP

$n_{i[-F]}$　　√TÍ　　　　　　　　$n_{i[+F]}$　　√TÍ
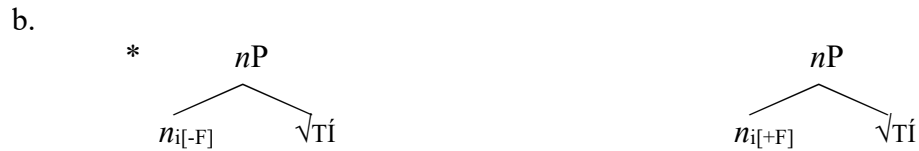
　　El tío de Ana y la <tía> de Clara llegaron.
　　Intended: 'Ana's uncle and Clara's (aunt) arrived.'

(repeated from (31)d)

In brief, the identity condition proposed in this dissertation thus far can straightforwardly account for the ill-formed status of mismatches with Class I nouns.

However, given that the identity condition is predicated on featural *non-distinctness*, the following question arises: in the case of feminine-masculine mismatches with Class I nouns, why can't there be a nominal in the ellipsis site whose √ROOT was licensed by default *n*? This configuration would satisfy the identity condition and be well-formed, contrary to fact. Additionally, masculine concord on the remnant determiner would be expected, since default grammatical gender is masculine.

The answer to why plain *n* could not be in the ellipsis site lies in the proposed licensing conditions for Class I. Consider again a √ROOT like √TÍ, which underlies 'uncle/aunt'. If

such a √ROOT arrives in the interpretive component in the context of plain *n*, a licensing violation occurs. In other words, √TÍ cannot be interpreted in such a structural context (77)c:

(77)   *The Encyclopedia and Class I: √TÍ*
       a. *Encyclopedia*   [*nP* [*n*i[+F] [√TÍ]]   is a person, who uses female pronouns…
       b. *Encyclopedia*   [*nP* [*n*i[-F] [√TÍ]]   is a person, who uses male pronouns…
       c. *Encyclopedia*   [*nP* [*n* [√TÍ]]   is ???

Since the issue that arises in combining a √ROOT like √TÍ with plain *n* is unrelated to externalization, ellipsis cannot repair the violation (see section 4.3.3):

(78)   *[GENDER] mismatch with Class I nouns (repeated from (31)h)*
       a. *La tía de Ana y el <[*nP* [***n*i[-F]** [√TÍ]] > de Clara llegaron.
                                                          violates featural non-distinctness
       b. *La tía de Ana y el <[*nP* [***n*** [√TÍ]] > de Clara llegaron.
                                          licensing violation; cannot be repaired by ellipsis

In sum, we have derived the ill-formed status of feminine-masculine Class I mismatches, whether the ellipsis site contain *n*i[-F] or plain *n*.

At this juncture, notice that Class I mismatches cannot adjudicate between the identity condition proposed here and others that require strict syntactic identity, given that a feature clash is ruled out under either approach. In fact, the analysis in Saab 2004: 94-95 of the ill-formed mismatch involving nouns like *tío/tía* 'uncle/aunt' is identical to the proposal here (for Saab though, (78)b would violate the identity condition, since plain *n* is not identical to *n*i[+F]). However, as we will see below, Class II mismatches *can* adjudicate between the two proposals.

Moving on, then, recall that Class II is the diametrical opposite of Class I in relation to [GENDER] mismatches—with Class II nouns, mismatches are allowed symmetrically. I propose that √ROOTs in this class are licensed either under *n*i[+F] or plain *n*. The latter

configuration yields a default, leading to masculine concord on all elements in the nominal domain:[113]

(79)   *Licensing conditions for Class II*
       a.   *n*P   *feminine*

             n[+F]   √II
       b.   *n*P   *default*

             n   √II
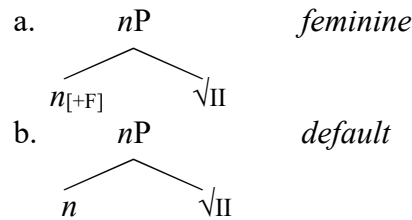
There exists some evidence that the grammatically masculine version of Class II noun pairs is licensed by plain *n*. Consider the contrast below:[114]

(80)   *Class II—masculine noun is default*
       a. Laura es **tía**   de Mario.
          Laura is **aun**t of Mario
          'Laura is Mario's aunt.'
       b. *Laura es **tío** de Mario.
          Intended: 'Laura is Mario's aunt.'
       c. Laura es **abogada**.
          Laura is **lawyer.F**
          'Laura is a lawyer.'
       d. ?/?? Laura es **abogado**.
          'Laura is a lawyer.'

---

[113] Whether √ROOTs in Class II can be licensed under $n_{i[-F]}$ is immaterial for the analysis, since a mismatch could always involve a default *n* in the antecedent (masculine-feminine mismatch) or ellipsis site (feminine-masculine mismatch).

[114] There exists microvariation regarding these judgments today, since speakers are particularly attuned to the manipulation here and hesitate to accept the grammatically masculine version, though the contrast with Class I nouns is sharp. It would be interesting to see whether the degree of acceptance for an example like (80)d correlates with a specific behavior of the nouns I have classified as Class II under ellipsis, as well as with an individual's awareness of issues related to gender identity.

The data above show that speakers accept (to different degrees) the use of a grammatically masculine noun denoting a profession (Class II) for an individual who uses female pronouns (see Harris 1991). The contrast with Class I nouns is sharp (see below on Class III as well). I interpret the facts above as independent evidence that the grammatically masculine form of noun pairs in Class II can be licensed by plain n.

Given my proposal about the structure of Class II nominals in (79), the pattern is derived as follows:

(81)   *[GENDER] mismatch with Class II satisfies featural non-distinctness*

a.   *Antecedent*   *Ellipsis site*
✔ nP nP

$n_{i[+F]}$ √ABOGAD n √ABOGAD

?La abogada de Ana es competente y el <abogado> de Eu también.
'Ana's (female) lawyer is competent and Eu's (male lawyer) is too.'
(repeated from (35)d)

b.
✔ nP nP

n √ABOGAD $n_{i[+F]}$ √ABOGAD

El abogado de Ana es competente y la <abogada> de Eu también.
'Ana's (male) lawyer is competent and Eu's (female lawyer) is too.'
(repeated from (34)d)

Under the proposal advanced here for the licensing conditions on Class II, the derivations above show how [GENDER] mismatches with these nouns never result in a featural clash; i.e., there is no configuration where a $n_{i[+F]}$ clashes with a $n_{i[-F]}$. Rather, the examples here involve a mismatch between a default n which has no [GENDER] feature, and a $n_{i[+F]}$. This mismatch configuration is ruled-in by the identity condition proposed in this dissertation.

Note that the well-formed status of mismatches with Class II nouns suggests that the current proposal is on the right track, when assessed in comparison to identity conditions that require strict identity. Meanwhile, let us consider how an identity condition like the one proposed in Rudin 2019 might address the data I have discussed here. One could start by cleaving apart the nominal spine (in analogy to the clausal spine) and demarcate a specific area that is taken into account by the identity condition, and another that is not taken into account by the identity condition. Let us assume that there is a natural separation between elements inside nP—let us call this area, the *nominal core*— and elements above it. We could assume that this demarcation captures the overall contrast between [NUMBER] and [GENDER] mismatches under ellipsis. However, problems arise immediately. First, under our analytical assumptions, [GENDER] *is* within the nominal core, and inside the ellipsis site in NP ellipsis and predicate ellipsis. If every element in the ellipsis site must find a structure-matching correlate in the antecedent, then we would predict that [GENDER] mismatches would be disallowed *across the board*, contrary to fact (recall our Class II mismatches above; see 4.4.2 below for Class III mismatches). For example, a plain n in the ellipsis site in (81)a would be unable to find a structure-matching correlate, leading to a violation of the condition in Rudin 2019, regardless of how we modify the original formulation of the condition to address DP internal mismatches. In brief, the empirical picture of [GENDER] mismatches cross-linguistically show that a condition like Rudin 2019's is untenable (see more in 4.7 for Saab 2010a's condition).[115]

---

[115] A conceivable alternative is one that would posit that there are two loci for [GENDER] in the nominal spine; e.g., for Class I pairs, [GENDER] would be on *n,* while for Class II pairs, it would be located higher (and thus

One brief final aside is necessary before I end our discussion of Class I and II nominals. As I have pointed out before, Class I nouns exhibit an alternating theme vowel; *-o* for the masculine and *-a* for the feminine. Some Class II nominals alternate in this way as well, showing that there is no morphological correlate that signals class membership in relation to the ellipsis facts. The literature on the ending of grammatically masculine versus feminine nouns in Spanish is vast, and has shown that there exist tendencies, but no absolutes, regarding which ending appears on nouns of different grammatical gender (see among many others, Harris 1991, Halle & Marantz 1994, Embick 2010). For instance, there are grammatically masculine nouns that end in *-a* (e.g., (*el*) *día* 'day'), grammatically feminine nouns that end in *-o* (e.g., (*la*) *mano* 'hand'), and many that display an *-e* (e.g. *padre/madre* 'father/mother') or no vowel at all (e.g. (*el*) *tapir* 'the tapir' and (*la*) *codorniz* 'the quail'). I refer the reader to Kramer 2015: 235-243 for an analysis of these nominal endings that is couched in the analytical assumptions of this chapter (cf. Bermúdez-Otero 2013). In a nutshell, Kramer posits that the arbitrary endings depend on membership in declension classes (that are different from the classes discussed in this work in relation to ellipsis). Kramer proposes that the vowel (or lack thereof) observed on nominals is the spell-out of a Theme node that is inserted post-syntactically and adjoins to *n*. The exponence of the feature borne by this Theme node is conditioned by the √ROOT or *n*; different features undergo Vocabulary Insertion and surface as the vowels *-o*, *-a*, *-e*, or a null morpheme. The crucial takeaway for our purposes is that there is no correlation between the appearance of a specific theme vowel on members of a noun pair and their

---

outside of the ellipsis site). This hypothetical alternative cannot account for the predicate ellipsis examples nor the asymmetric behavior of Class III pairs, so I set it aside.

behavior under ellipsis. Furthermore, this lack of correlation between declension classes and ellipsis-related classes makes the acquisition profile difficult and might also underlie the degree of inter-speaker variation observed (see section 4.7).

To summarize this sub-section, I first proposed a set of licensing conditions for Class I and II nouns. I then showed how the identity condition proposed in this dissertation accounts for the pattern of [GENDER] mismatches with nouns in both of these classes. I also demonstrated that certain derivations that do comply with featural non-distinctness are nevertheless ruled-out by the independent proposal that certain gaps—those whose source is semantic—cannot be repaired by ellipsis. Finally, I noted the lack of correlation between the morphological cue of theme vowels and noun class membership.

4.4.2 Class III mismatches

We are now in a position to address the asymmetrical behavior of Class III nouns. The core puzzle that must be accounted for is why [GENDER] mismatches are well-formed only when the grammatically masculine member of a pair is in the antecedent.

This subsection is organized as follows: first, I will show how the distinction between violations that are repairable by ellipsis and those that are not repairable by ellipsis accounts for the well-formed masculine-feminine mismatches with Class III nouns:

(82)   *Class III and [GENDER] mismatches: masculine-feminine*
       A mismatch where the masculine is in the antecedent and the feminine is in the
       ellipsis site is allowed because of repair-by-ellipsis of a morphophonological gap.

The well-formed status of masculine-feminine mismatches is only half of the empirical puzzle, though—we must also explain the ill-formed feminine-masculine mismatches. Before doing so, I will establish another ingredient of my overall proposal, laying out independent empirical support for it. I will present evidence that the syntactic identity

273

condition on ellipsis necessitates a second statement, in addition to the requirement proposed thus far that the antecedent and ellipsis site be featurally non-distinct. This second statement requires that √ROOTs in the antecedent and ellipsis site be identical (see Saab 2008). Having proposed a final identity condition combining the two statements—one related to features, the other related to √ROOTs—I will return to the ill-formed feminine-masculine mismatches. I will propose that the additional morpheme on the feminine Class III nouns is a √ROOT. I will show, thus, how this mismatch configuration is correctly ruled out by the final identity condition proposed here.

Let us start with the first half of the empirical puzzle. Recall that [GENDER] mismatches with Class III nouns are allowed when the masculine version of a noun is in the antecedent—this same pattern is replicated in Greek, Brazilian Portuguese, and Russian (see 4.2):

(83)   *Class III [GENDER] mismatch—masculine-feminine (modified from (39))*
      a. El   actor de Hollywood ganó un Óscar y   la   actr-iz de
        the.M actor of Hollywood won an Oscar and the.F actor-IZ of
        Bollywood también.
        Bollywood also
        'The Hollywood actor won an Óscar and the Bollywood actress did too.'
      b. ?El actor de Hollywood ganó un Oscar y la *<ellipsis site>* de Bollywood también.
        'The Hollywood actor won an Oscar and the Bollywood (actress) did too.'

(84)   *Class III [GENDER] mismatch—masculine-feminine*
      a. Joel no   es alcalde   de Mixco, pero Andrea sí   es alcalde-sa de Mixco.
        Joel NEG is mayor.M of Mixco but   Andrea yes is mayor.SA of Mixco
        'Joel is not the mayor of Mixco, but Andrea is the mayor of Mixco.'
      b. ?Joel no es alcalde de Mixco, pero Andrea sí *<ellipsis site>*.
        'Joel is not the mayor of Mixco, but Andrea is.'

My proposal is the following: first, the masculine version of nouns in Class III contains a √ROOT that is licensed by plain *n*:

(85) *Licensing conditions for Class III—masculine nouns*

          *n*P   e.g., *actor, alcalde, héroe*

          /\

     n   √III
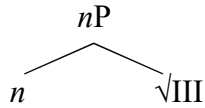
Just like with Class II nouns, there is some evidence that the grammatically masculine version of Class III noun pairs is licensed by plain *n*, and not $n_{i[-F]}$. There is a contrast between Class I and Class III nominals when the grammatically masculine member of a noun pair is used predicatively (the same caveats expressed in footnote 114 hold here):
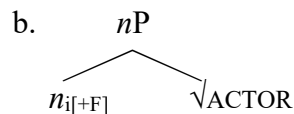
(86) *Class III—masculine noun is default*

    a. Laura es **tía**   de Mario.
       Laura is **aun**t of Mario
       'Laura is Mario's aunt.'

    b. *Laura es **tío** de Mario.
       Intended: 'Laura is Mario's aunt.'

    c. Laura es **alcaldesa** de Mixco.
       Laura is **mayor.F** of Mixco
       'Laura is mayor of Mixco.'

    d. ?/?? Laura es **alcalde** de Mixco.
       'Laura is mayor of Mixco.'

My proposal for our crucial examples is that the ellipsis sites in (83)b and (84)b do *not* contain the grammatically feminine counterparts of the Class III nouns pairs; e.g., (83)b does not contain the pronounceable counterpart of *actor*, namely *actriz*. Rather, the ellipsis site contains a nominal composed of the same √ROOT as *actor* (notated √ACTOR below), albeit licensed by $n_{i[+F]}$ (85)b:

(87) *Licensing conditions for Class III—*√ACTOR

    a.   *n*P   *actor*

        /\

     n   √ACTOR

    b.   *n*P   *no Vocabulary Insertion*

        /\

   $n_{i[+F]}$ √ACTOR

I propose that the combination of a root like $\sqrt{\text{ACTOR}}$ with $n_{\text{i}[+\text{F}]}$ leads to an insertion problem in the externalization component. In other words, there is a morphophonological gap (which is repairable by ellipsis):

(88)   *Class III nouns: morphophonological gaps*
         Vocabulary Insertion $\sqrt{\text{ACTOR}} \leftrightarrow$??? / $n_{\text{i}[+\text{F}]}$ _
         Encyclopedia $[_{n\text{P}}\ [n_{\text{i}[+\text{F}]}\ [\sqrt{\text{ACTOR}}]]$ is interpreted as a performer, who uses female pronouns…

In other words, I propose that the grammar generates two different expressions whose encyclopedic insertion is 'actor that uses female pronouns'. However, one of these expressions is unpronounceable due to a morphological gap—i.e., there is no appropriate Vocabulary Insertion rule (88) (recall our discussion of gaps in the paradigm for Spanish *abolir* 'to abolish' in (63) above).

Putting it all together, I propose that well-formed masculine-feminine mismatches with Class III nouns satisfy featural non-distinctness and involve repair; the masculine in the antecedent is licensed by plain *n*, while there is a $n_{\text{i}[+\text{F}]}$ in the ellipsis site:

(89)   *[GENDER] mismatch with Class III nouns: masculine-feminine*
         ?El **actor** de Hollywood ganó un Oscar y la <$[_{n\text{P}}\ [\boldsymbol{n}_{\textbf{i}[+\textbf{F}]}\ [\sqrt{\text{ACTOR}}]]$> de Bollywood también.
         'The Hollywood actor won an Oscar and the Bollywood (actress) did too.'
                                    satisfies featural non-distinctness; repair-by-ellipsis

We have thus derived the well-formed mismatches with Class III nouns where the antecedent contains a masculine noun. The question at this juncture is why the inverse configuration is impossible—i.e., why can't feminine nouns like *actriz* 'actress', *alcaldesa* '(female) mayor', or *heroína* '(female) hero' be in the antecedent?

My proposal for deriving the ill-formed status of feminine-masculine mismatches incorporates a second ingredient with independent support in the literature. Concretely, I

276

propose that the overall syntactic identity condition on ellipsis is composed of *two* statements. The first statement is our now-familiar requirement that *features* satisfy non-distinctness. The second statement is not about features, but is specific to √ROOTs:

(90)   *Identity condition on √ROOTs (repeated from (1))*
There must be a strict one-to-one match between all √ROOTs properly contained in the ellipsis site and √ROOTs in the antecedent.

Recall that I take √ROOTs to be featureless syntactic primitives that are individuated via an index that provides instructions to the post-syntactic components of the grammar. Given this, I do not think that it should worry us from a conceptual standpoint that the syntactic identity condition on ellipsis is divided into two statements, since each statement is a requirement on *different primitives*—features on the one hand, √ROOTs on the other.

The condition above is directly inspired by Saab 2008 (which we return to in 4.6) and does more work than just deriving the ill-formed status of Class III feminine-masculine mismatches, as I will show below. For example, it can capture well-known facts about lexical identity under ellipsis (Rooth 1992, Chung 2006): e.g., the sluicing example in (91)b cannot be interpreted as intended—even though the act of slapping someone might entail that a humiliation took place—since the verbs *slap* and *humiliate* each contain different √ROOTs; (91)d, meanwhile, is ill-formed because *something* inside the ellipsis site contains a √ROOT that is not present in the antecedent, violating (90):

(91)   *√ROOT identity under ellipsis (interpretation unavailable)*
a. Ali slapped Foreman, but I can't remember who else <Ali slapped>.
b. *Ali **slapped** Foreman, but I can't remember who else <Ali **humiliated**>.
                                                                 (Deniz Rudin p.c.)
c. She read **something**, but we're not sure by which author$_1$ <she read **something** t$_1$>.
d. *She read, but we're not sure by which author$_1$ <she read **something** t$_1$ >.
                                                                 (Chung 2006)

Other data provide strong support as well for a condition like (90). First, recall that [NUMBER] can mismatch under ellipsis:

(92)   *Number mismatch allowed*
    a. Bob's wig is on the desk and Monet's wigs are under the desk.
    b. Bob's wig is on the desk and Monet's <wigs> are under the desk.   ✓SG-PL
    c. Bob's wigs are on the desk and Monet's wig is on the desk.
    d. Bob's wigs are on the desk and Monet's <wig> is on the desk.   ✓PL-SG

With this in mind, consider the data below, where a mismatch is forced between two lexical items that are synonymous but differ in the number agreement they trigger on the remnant in the target clause (Merchant 2019; see discussion in Saab 2019): *wedding* and *nuptials*.[116] The mismatch is forced via a manipulation of the remnant:[117]

---

[116] Saab 2019 offers parallel Spanish examples where a mismatch is forced between the masculine *casamiento* and the feminine *boda* (both being synonymous and meaning 'wedding'). However, it is impossible to know whether the effect we see is due to the violation of √ROOT identity only, or whether the effect is due to the violation of √ROOT identity *and* a clash in [GENDER] features. Those Spanish examples thus do not provide evidence for a √ROOT identity requirement (see section 4.5.4 on Class VIII nouns—namely, inanimate nouns that are compatible with masculine or feminine determiners but cannot mismatch in [GENDER] under ellipsis; e.g., *el/la mar* 'the sea').

[117] The example in (93)e should be well-formed under a reading where the elided nominal is *weddings*, as opposed to *nuptials*; i.e., when the intended interpretation is one where Rachel has celebrated several events at Rockefeller Chapel. While neither Merchant nor Saab indicate if this is indeed the case, speakers I have consulted confirm that this reading is available, as expected. Furthermore, we must assume that the ill-formed status of the singular *nuptial* is the result of an encyclopedic gap (i.e., it cannot be repaired by ellipsis). If *nuptial* were ill-formed due to a morphological gap, we would expect (93)f to be possible with an ellipsis site containing *nuptial*, due to repair-by-ellipsis.

(93)   *√ROOT identity under ellipsis (mismatch forced via the remnant)*
   a. Beth's wedding was in Bond Chapel, and Rachel's wedding was in Rockefeller chapel.
   b. Beth's wedding was in Bond Chapel, and Rachel's <wedding> was in Rockefeller chapel.
   c. Beth's nuptials were in Bond Chapel, and Rachel's nuptials were in Rockefeller chapel.
   d. Beth's nuptials were in Bond Chapel, and Rachel's <nuptials> were in Rockefeller chapel.
   e. *Beth's **wedding** was in Bond Chapel, and Rachel's <**nuptials**> **were** in Rockefeller chapel.
   f. *Beth's **nuptials** were in Bond Chapel, and Rachel's <**wedding**> **was** in Rockefeller chapel.

As we observed, [NUMBER] mismatches are well-formed under ellipsis, so the manipulation above should be well-formed *if a √ROOT mismatch were possible*. However, such a mismatch is ill-formed. We therefore have evidence that there is a condition that pertains to √ROOTs, which are distinct syntactic formatives from features. We will return to some other consequences of the √ROOT condition in chapter 5, where we assess Chung's generalization (Chung 2006).

I will assume in what follows, then, that the syntactic identity condition on ellipsis is divided into two statements, encapsulated in the final condition proposed in this dissertation, given below:

(94)   *Syntactic identity condition on ellipsis (final)*
   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.

Now that I have established this final identity condition, we are ready to return to the pattern of [GENDER] mismatches that arises with Class III nominals.

Recall that that [GENDER] mismatches with Class III nouns display an asymmetrical pattern—a mismatch is well-formed only if the masculine noun is in the antecedent. Furthermore, I noted a morphological trait of feminine nouns in this class—they bear an additional morpheme; e.g., *actr-iz* 'actress', *alcalde-sa* 'female mayor', *hero-ína* 'female hero':

(95) *Class III [GENDER] mismatches: feminine-masculine (modified from (40))*
    a. La   actr-iz   de Bollywood ganó un Óscar y   el   actor de
       the.F actor-IZ of Bollywood won   an Oscar and the.M actor of
       Hollywood también.
       Hollywood also
       'The Bollywood actress won an Oscar and the Hollywood actor did too.'
    b. *La actriz de Bollywood ganó un Óscar y el *<ellipsis site>* de Hollywood
       también.
       Intended: 'The Bollywood actress won an Oscar and the Hollywood (actor) did
           too.'

(96) *Class III [GENDER] mismatches: feminine-masculine*
    a. Andrea no es alcalde-sa de Mixco, pero Joel sí es alcalde de Mixco.
       Andrea not is mayor-SA of Mixco but   Joel yes is mayor of Mixco
       'Andrea is not the mayor of Mixco, but Joel is the mayor of Mixco.'
    b. *Andrea no es alcaldesa de Mixco, pero Joel sí *<ellipsis site>*.
       Intended: 'Andrea is not the mayor of Mixco, but Joel is.'

I build on Alexiadou 2017, which, in discussing Class III ellipsis data from Greek that are parallel to the Spanish data here, proposes that the additional suffix on the feminine noun (called "derivational" in that paper) is to blame for the asymmetric behavior of this class.[118]

___

[118] See also Bobaljik & Zocca 2011 and Polinsky (2020), the latter who notes that Class III feminine nouns bear derivational affixes that distinguish them from the masculine. The proposal here for Class III could be directly applied to explain the asymmetrical behavior of Class III in Brazilian Portuguese, Greek and, Russian. Interestingly, Donatelli 2019: 194-195 rejects this line of analysis for Spanish, claiming that nouns like *actor / actriz* are morphologically dissimilar enough that such a proposal can be set aside.

I propose that the suffix that appears on a Class III feminine noun is not an affix, but a √ROOT—i.e., the *-iz* portion in *actriz* 'actress', the *-sa* portion in *alcaldesa* '(female) mayor', the *-ína* portion in *heroína* 'female hero', etc., are √ROOTs that cannot surface independently, but must combine with another √ROOT.[119] My proposal follows recent work that argues that some affixes are not the spell-out of functional heads, but √ROOTs (Creemers et al. 2018; see Loewenstamm 2015 for the proposal that *all* affixes are √ROOTs).[120]

(97)   *Licensing condition for Class III feminine nouns*

nP   actriz, alcaldesa, heroína

$n_{i[+F]}$ √

√III   √IZ / √ESA / √ÍNA

---
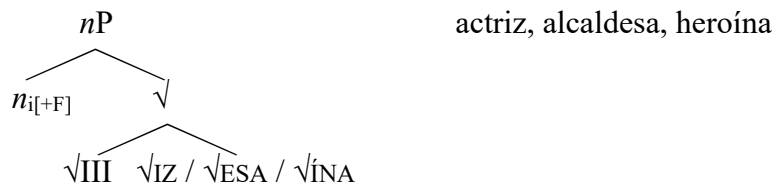
[119] My proposal is that these morphemes in Spanish are akin to English cran-morphs, which cannot be freestanding. Creemers et al. (2018) (see below in the main text) encode *featurally* the selectivity of the affixes they analyze as √ROOTs—for these authors, some √ROOTs are endowed with an uninterpretable √ feature that must be checked by merging with another √ROOT. It is unclear to me that this is the only way to limit the distribution of the relevant morphemes. Additionally, if one follows Harley 2014, endowing √ROOTs with features creates problems for their status as syntactic primitives that differ from features (see also Saab 2010a, who endows √ROOTs with [GENDER] features.

[120] I leave for the future an investigation of whether the morphemes that appear with Class III feminine nouns display the category flexibility that Creemers et al. attribute to all affixes argued to be √ROOTs. Some preliminary evidence suggests this may be on the right track. Consider *-sa*, which surfaces attached to adjectives as well—e.g., *torpe* 'clumsy', (*la*) *torpeza* 'clumsiness'; *áspero* 'rough', (*la*) *aspereza* 'roughness' (note that the orthographic <z> in Latin American Spanish is pronounced [s]).

Taking *actriz* as representative of Class III feminine nouns, observe below my proposal for its full set of licensing conditions:

(98)   *Licensing conditions for Class III actriz 'actress'*

a.   *n*P   actriz

$n_{i[+F]}$ √
        √ACTOR   √IZ

b.*   *n*P   *Intended*: (male) actriz

$n_{i[-F]}$ √
        √ACTOR   √IZ

c.*   *n*P   *Intended*: (default) actriz

n   √
        √ACTOR   √IZ
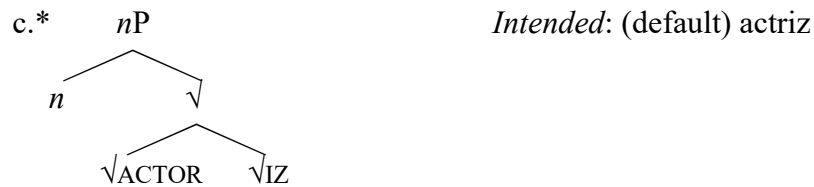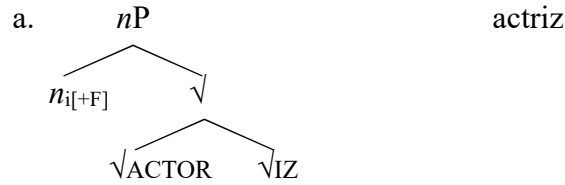
Given the proposal above, we can now go over why a feminine-masculine [GENDER] mismatch is impossible with Class III nouns. In a configuration where a noun like *actriz* is in the antecedent, a [GENDER] mismatch forces a nominal in the ellipsis site that is licensed by $n_{i[-F]}$ or plain *n*. These configurations give rise to licensing violations that cannot be repaired by ellipsis:

(99)   *Class III [GENDER] mismatch: feminine-masculine*

a.   *   *Antecedent*   *Ellipsis site*
                  *n*P   *n*P

        $n_{i[+F]}$ √ $n_{i[-F]}$ √
                √ACTOR   √IZ   √ACTOR   √IZ

b. * *Antecedent*   *Ellipsis site*

```
        nP                              nP
       /  \                            /  \
   n_i[+F]  √                       n     √
           /  \                          /  \
      √ACTOR  √IZ                   √ACTOR  √IZ
```
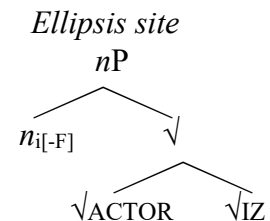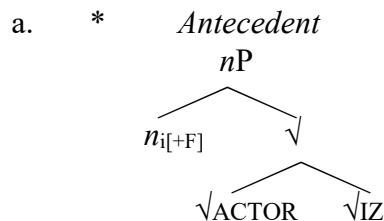
Our job is not done, however. The following must also be ruled out: a derivation of feminine-masculine mismatches with Class III nouns where the ellipsis site contains a √ROOT like √ACTOR licensed by $n_{i[-F]}$. This derivation is ill-formed as a result of a violation of √ROOT identity (94)b, since the √ROOTs in the antecedent and ellipsis site do not match one-to-one:

(100) *Class III [GENDER] mismatch: feminine-masculine*

     * *Antecedent*   *Ellipsis site*

```
        nP                              nP
       /  \                            /  \
   n_i[+F]  √                       n    √ACTOR
           /  \
      √ACTOR  √IZ
```
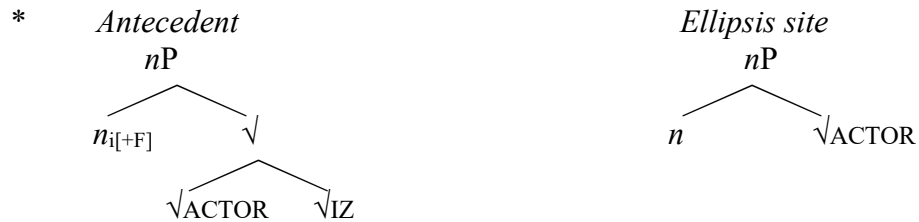
Finally, we can now circle back for completeness to the well-formed mismatches with Class III nouns where the masculine is in the antecedent. Recall that I proposed that in this example type, a noun like *actriz* is not in the ellipsis site (see (89)). Such a configuration is also ruled out as a violation of √ROOT identity (94)b:

(101) *Class III [GENDER] mismatch: masculine-feminine (ill-formed derivation)*

     Antecedent   *Ellipsis site*

```
   *    nP                              nP
       /  \                            /  \
      n   √ACTOR                   n_i[+F]  √
                                           /  \
                                      √ACTOR  √IZ
```

Instead, what the ellipsis site contains in well-formed masculine-feminine mismatches is the root √ACTOR directly licensed by $n_{i[+F]}$, an expression that suffers a morphological gap

which, as expected, is repaired under ellipsis (89). The behavior of [GENDER] mismatches with Class III nouns under ellipsis has now been derived.

To summarize this sub-section, I began by describing the overall puzzle raised by Class III mismatches: only a masculine noun is well-formed in the antecedent, not a feminine noun. I first proposed that well-formed masculine-feminine mismatches result from repair-by-ellipsis. I then proposed, and independently justified, a second portion of the identity condition that is specific to √ROOTs and requires strict one-to-one matching between antecedent and ellipsis site:

(102)   *Syntactic identity condition on ellipsis (final)*
    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.

I ended by illustrating how this final identity condition can account for the ill-formed status of feminine-masculine mismatches with Class III nouns.

4.5 Expanding the empirical picture

In this section, I expand the empirical picture that was presented in section 4.2 and discuss a larger sample of Spanish nouns and their [GENDER] mismatch patterns under ellipsis. I then show how the analysis developed in the last section can account for the data here as well.

My contribution is twofold. First, I intend to demonstrate that the tools employed in 4.3 to analyze the three analytically significant classes of nouns can be extended to account for the patterning under ellipsis of a broader sample of lexical items. As we will observe, for instance, the proposal from 4.3.3 that certain gaps are semantic in nature plays a key role in deriving the ill-formed status of several derivations that I will discuss. In a nutshell, my

goal is to collect a set of data that could be perceived as exceptions to the generalizations established before and show that the data are not problematic for the proposed analysis. Second, I hope that the analytical conclusions stemming from the discussion of these data will encourage researchers to delve into similar explorations in other languages.

A comment is in order before I begin the description. None of the classes of nouns described in what follows constitute a "class" in a way that is analytically parallel to the way in which Classes I, II, and III constitute classes. In other words, Class I and Class II are distinguished from each other because they each exhibit a different [GENDER] mismatch pattern; the same can be said for the distinction between Class I and III, and the distinction between Class II and III. In contrast, my taxonomy of remaining classes (Classes IV – XI) is not meant to be indicative of hitherto undiscussed patterns of [GENDER] mismatch under ellipsis; i.e., it is not the case that Class IV allows [GENDER] mismatches only when a feminine noun is in the antecedent[121], Class V allows some other pattern, and so on. Rather, these groupings represent one way to slice a portion of the Spanish lexical pie into natural classes, based on grammatical or semantic criteria, in order to demonstrate how the analysis developed in the prior section can be applied productively to derive ellipsis data using a broader sample of lexical items.

Some of the nouns I discuss in this section had been discussed previously in the ellipsis literature (occasionally under a different labelling from the one proposed here; see my comments on the classification of Class IV and V in 4.5.1). As I will show, mismatches going both ways (i.e., masculine antecedent-feminine ellipsis site; feminine antecedent-

---

[121] To my knowledge, this kind of asymmetrical pattern has never been documented in the literature.

masculine ellipsis site) cannot be tested in several cases, since there exist gaps in certain

noun pairs for the grammatically masculine or feminine version of the noun.

Naturally, the categorization here could be rethought and refined. I have attempted,

however, to group nouns given certain characteristics that allows us to discuss them in an

orderly manner that might entice further cross-linguistic research. For example, Class IV

is constituted by grammatically feminine nouns for which there is no grammatically

masculine counterpart sharing the same underlying √ROOT; Class VIII is composed of

inanimate nouns, etc.

With this in mind, a summary of the data I will assess is below:

(103)   *[GENDER] mismatches under ellipsis—different classes of nouns*
    a. *Class IV*: feminine – masculine mismatch ill-formed   e.g., *nuera*
    b. *Class V*:   masculine – feminine mismatch ill-formed   e.g., *yerno*
    c. *Class VI*: feminine – masculine mismatch ill-formed   e.g., *institutriz*
    d. *Class VII*: masculine – feminine mismatch ill-formed   e.g., *músico*
    e. *Class VIII*: mismatch symmetrically ill-formed   e.g., *orden*
    f. *Class IX*: feminine – masculine mismatch ill-formed   e.g., *víctima*
    g. *Class X*:   feminine – masculine mismatch ill-formed   e.g., *serpiente*
    h. *Class XI:* masculine – feminine mismatch ill-formed   e.g., *delfín*

4.5.1 Classes IV and V

Let us start our discussion by delving into two classes of nouns for which there is a gap in

either the masculine (Class IV) or feminine cell (Class V) of a (potential) noun pair (see

below for an approach that would collapse nouns in these classes together).

First, consider a sample of nouns belonging to what I am calling Class IV. This class

is composed of grammatically feminine nouns:

(104)   *Class IV nouns*
    a. vaca   'cow'
    b. iguana   'female iguana'
    c. yegua   'mare'
    d. nuera   'daughter-in-law'

e. madre   'mother'

Nouns in this class denote (i) individuals who use feminine pronouns or (ii) female animals.

Moving on to ellipsis, observe that a [GENDER] mismatch with Class IV nouns is ill-formed, as illustrated below with NP ellipsis:

(105)   *Class IV and NP ellipsis: [GENDER] mismatch ill-formed*
    a. La   vaca de Aída es torpe   y   la   vaca de María también.
      the.F cow of Aida is clumsy and the.F cow of María too
      'Aída's cow is clumsy and María's cow is too.'

    b. La vaca de Aída es torpe y la <vaca> de María también.
      'Aída's cow is clumsy and María's (cow) is too.'

    c. *La vaca de Aída es torpe y el <*ellipsis site*> de María también.
      Intended: 'Aída's cow is clumsy and María's (male cow) is too.'

Ellipsis of a larger portion of structure gives rise to the same pattern; here, I use our familiar manipulation of a proper name in the remnant:

(106)   *Class IV and predicate ellipsis: [GENDER] mismatch ill-formed*

    a. Aída no es nuera   de Pedro pero María sí   es nuera
      Aída not is daughter.in.law of Pedro but   María yes is daughter.in.law
      de Pedro
      of Pedro
      'Aída is not Pedro's daughter-in-law, but María is Pedro's daughter-in-law.'

    b. Aída no es nuera de Pedro pero María sí <es nuera de Pedro>.
      'Aída is not Pedro's daughter-in-law, but Maria is (Pedro's daughter-in-law).'

    c. *Aída no es nuera de Pedro pero Sebas sí <*ellipsis site*>.
      Intended: 'Aída is not Pedro's daughter-in-law, but Sebas is (Pedro's son-in-law).'

My interpretation of these data is that nominals in Class IV are composed of √ROOTs that are licensed only by $n_{i[+F]}$, as shown in (107)a below. The √ROOTs in Class IV nouns cannot, however, be licensed by $n_{i[-F]}$ (107)b or plain n (107)c. In other words, the gap in the masculine cell for Class IV is *not* the result of a lack of morphophonological exponent (i.e., the issue here is *not* parallel to combining a √III with $n_{i[+F]}$; see 4.3). Rather, the gap is the result of a licensing violation, which I argued cannot be repaired by ellipsis (see the

feminine-(default) masculine mismatch with Class II and the feminine-masculine mismatch with Class III in section 4.4):

(107) *Licensing conditions for Class IV nominals*

    a.   *n*P   feminine

        $n_{i[+F]}$ √IV

    b. *   *n*P   *Intended*: masculine

        $n_{i[-F]}$ √IV

    c. *   *n*P   *Intended*: default

        n   √IV

The proposal in (107) accounts for the ill-formed status of examples like (105)c and (106)c: the mismatch forces a Class IV noun inside the ellipsis site whose √ROOT is licensed either by $n_{i[-F]}$ or bare *n* (either of which would lead to masculine concord on the remnant in NP ellipsis), a licensing violation in the sense discussed in 4.3.2. In short, the structure forced inside the ellipsis site cannot be repaired by ellipsis.

Consider now Class V. In this class, only a grammatically masculine noun is available and there is a gap in the feminine cell. Nouns in this class denote either (i) individuals who use masculine pronouns or (ii) male animals. Notice that the sample below consists of the semantic counterparts of the Class IV nouns in (104):

(108) *Class V nouns: only masculine noun exists*
    a. toro   'bull'
    b. garrobo   'male iguana'
    c. caballo   'horse'
    d. yerno   'son-in-law'
    e. padre   'father'

A [GENDER] mismatch with Class V nouns is also ill-formed:

(109)   *Class V and NP ellipsis [GENDER] mismatch ill-formed* [122]

   a. El   toro de Aída es torpe y   el   toro de María también.
      the.M bull of Aida is clumsy and the.M bull of María too
      'Aída's bull is clumsy and María's bull is too.'

   b. El toro de Aída es torpe y el \<toro\> de María también.
      'Aída's bull is clumsy and María's is too.'

   c. *El toro de Aída es torpe y la \<*ellipsis site*\> de María también.
      Intended: 'Aída's bull is clumsy and María's (female bull) is too.'

(110)   *Class V and predicate ellipsis: [GENDER] mismatch ill-formed*

   a. Sebas no es yerno   de Pedro pero Pablo sí   es yerno   de Pedro.
      Sebas not is son.in.law of Pedro but   Pablo yes is son.in.law of Pedro
      'Sebas is not Pedro's son-in-law, but Pablo is Pedro's son-in-law.'

   b. Sebas no es yerno de Pedro pero Pablo sí \<es yerno de Pedro\>.
      'Sebas is not Pedro's son-in-law, but Pablo is (Pedro's son-in-law).'

   c. *Sebas no es yerno de Pedro pero Aída sí \<*ellipsis site*\>.
      Intended: 'Sebas is not Pedro's son-in-law, but Aída is (Pedro's daughter-in-
      law).'

My interpretation of these data is similar to the analysis I proposed for ill-formed

mismatches with Class IV nouns. Specifically, I propose that Class V nouns contain a

√ROOT that is licensed only by one of the *n* heads discussed so far—namely, $n_{i[-F]}$ (111)b.

These √ROOTs cannot be licensed by $n_{i[+F]}$ (111)a or *n* (111)c.

(111)   *Licensing conditions for Class V nominals*

   a. *   *n*P   *Intended*: feminine

         $n_{i[+F]}$ √V

   b.   *n*P   masculine

         $n_{i[-F]}$ √V

---

[122] As Juan Uriagereka points out to me, some populations of Spanish speakers might have quite complex

representations for animals of different sexes, and these speakers might not have a gap for certain Class IV

or V nouns. For instance, some speakers do allow for a noun *tora*—a grammatically feminine counterpart of

the noun *toro* 'bull'— which denotes a cow that is in heat.

c. * nP   *Intended*: default

n √V
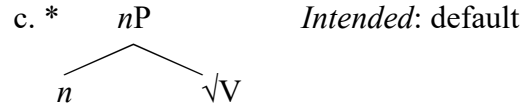
The ellipsis pattern we observe with Class V nouns follows from the licensing violation incurred by forcing a [GENDER] mismatch—ellipsis cannot repair a structure wherein a √ROOT in a Class V noun combines with $n_{i[+F]}$.

At this juncture, notice that Class IV and V nouns could be collapsed together into a single class, giving rise to a putative alternative classification involving semantically related pairs (e.g., *caballo/yegua* 'horse/mare'). It is necessary, then, to correctly rule out derivations where there exists a mix-and-(mis)match of a nominal from Class IV with a nominal from Class V under ellipsis.[123]

The final identity condition proposed in this dissertation rules out this configuration, since *both* statements in the identity condition are violated—(i) there is a clash in features, violating featural non-distinctness and (ii) there is no one-to-one match of √ROOTs, violating √ROOT identity. In essence, this is the same analysis as in Saab 2004, 2008, 2010a, albeit Saab assumes a strict identity condition on features—a condition which I have argued

---

[123] Yet another alternative classification of nominals in Class IV and V would take the relevant pairs to be composed of the *same* √ROOT and the different surface forms to arise as the result of √ROOT suppletion (i.e., a √ROOT is inserted as /*madre*/ in the context of $n_{i[+F]}$ and /*padre*/ elsewhere; see Kramer 2015: 94-95 for a tentative proposal along these lines). That would rule out the mismatches (as long as the relevant √ROOTs cannot be licensed by default *n*; see the licensing conditions above). For a debate on whether √ROOT suppletion exists, I refer the reader to Harley 2014 and Borer 2014, and references therein. Saab 2010a labels noun pairs like *padre/madre* ('father/mother') "suppletive" pairs, but the analysis takes each of the pair members to contain a different √ROOT (see Saab 2010a: example (39)).

in this dissertation undergenerates (see chapter 2, 3, and the analysis of Class II and III

mismatches in 4.4):[124]

(112) *No mix-and-match for Class IV and V—NP ellipsis*
  a. La   vaca de Aída es torpe y   el   toro de María también.
     the.F cow of Aída is clumsy and the.M bull of María too
     'Aída's cow is clumsy and María's bull is too.'

  b. *La vaca de Aída es torpe y el <toro> de María también.
     Intended: 'Aída's cow is clumsy and María's (bull) is too.'

  c. El   toro de María es torpe y   la   vaca de Aída también.
     the.M bull of María is clumsy and the.F cow of Aída too
     'María's bull is clumsy and Aída's cow is too.'

  d. *El toro de María es torpe y la <vaca> de Aída también.
      Intended: 'María's bull is clumsy and Aída's (cow) is too.'

(113) *No mix-and-match for Class IV and V—predicate ellipsis*
  a. Aída no es nuera   de Pedro, pero Sebas sí   es yerno   de Pedro.
     Aída not is daughter.in.law of Pedro but   Sebas yes is son.in.law of Pedro
     'Aída is not Pedro's daughter-in-law, but Sebas is Pedro's son-in-law.'

  b. *Aída no es nuera de Pedro, pero Sebas sí <es yerno de Pedro>.
     Intended: 'Aída is not Pedro's daughter-in-law, but Sebas is (Pedro's son-in-law).'

  c. Sebas no es yerno   de Pedro, pero Aída sí es nuera   de Pedro.
     Sebas not is son.in.law of Pedro but   Aída yes is daughter.in.law of Pedro
     'Sebas is not Pedro's son-in-law, but Aída is Pedro's daughter-in-law.'

  d. *Sebas no es yerno de Pedro, pero Aída sí <es nuera de Pedro>.
     Intended: 'Sebas is not Pedro's son-in-law, but Aída is (Pedro's daughter-in-law).'

---

[124] Saab 2010a claims that these mix-and-(mis)match data are even worse than Class I mismatches (e.g.,

tío/tía 'uncle/aunt' mismatches). Saab argues that the reason for the greater unacceptability of mismatching

padre/madre is because such a configuration violates the condition on features (which is strict for him) and

the condition on √ROOTs, whereas a *tío/tía* mismatch only violates the condition on features and satisfies the

condition on √ROOTs. I share Saab's relative judgement here (thank you to Howard Lasnik for raising this

question).

In a nutshell, the identity condition proposed in this dissertation correctly rules out the ill-formed status of the mix-and-(mis)match examples above.

To summarize this subsection, I discussed Class IV and Class V nouns. In each of these classes, there is a gap in the masculine (Class IV) or feminine (Class V) cells of a (potential) noun pair. We observed that a [GENDER] mismatch with nouns in these classes is impossible. I proposed licensing conditions for the nouns in both classes and showed how that proposal derives the observed pattern. I then showed how a mix-and-(mis)match derivation of nouns in these classes violates the identity condition two-fold: both (i) featural non-distinctness and (ii) strict √ROOT identity are violated.

4.5.2 Class VI

Class VI is composed of a single grammatically feminine noun:

(114)   *Class VI noun*
         institutriz   'governess'

The denotation of the noun above includes a specification that the individual uses feminine pronouns. For my consultants and I, there is no grammatically masculine counterpart of the noun which could be used for an individual who uses masculine pronouns:[125]

(115)   a. *Luis es institutriz.
             Luis is governess
             Intended: 'Luis is a governess.'
        b. *Luis es institutriz-o.
             Luis is governess-O
             Intended: 'Luis is a governess.'

---

[125] The dictionary entry for the *Real Academia Española* does not offer a male counterpart (see https://dle.rae.es/institutriz?m=30_2).

As shown below, a [GENDER] mismatch is ill-formed with the noun in this class:[126]

(116) *Class VI: no mismatch under NP ellipsis*

    a. Llegaron la   institutriz de Ana y   la   institutriz de Elsa.
       arrived   the.F governess of Ana and the.F governess of Elsa
       'Ana's governess and Elsa's governess arrived.'

    b. Llegaron la institutriz de Ana y la \<institutriz\> de Elsa.
       'Ana's governess and Elsa's arrived.'

    c. *Llegaron la institutriz de Ana y el \<*ellipsis site*\> de Elsa.
       Intended: 'Ana's governess and Elsa's (male governess) arrived.'

(117) *Class VI: no mismatch under predicate ellipsis*

    a. Mely no es institutriz, pero Fabiola sí   es institutriz.
       Mely not is governess but Fabiola yes is governess
       'Mely is not a governess, but Fabiola is a governess.'

    b. Mely no es institutriz, pero Fabiola sí \<es institutriz\>.
       'Mely is not a governess, but Fabiola is (a governess).'

    c. *Mely no es institutriz, pero Chepe sí \<*ellipsis site*\>.
       Intended: 'Mely is not a governess, but Chepe is (a male governess).'

I propose that the facts above suggest that the noun in this class is composed of a √ROOT

that can only be licensed by $n_{i[+F]}$:

(118) *Licensing conditions for Class VI nominal*

    a.   *n*P   feminine

          $n_{i[+F]}$ √VI

    b. *   *n*P   *Intended*: masculine

          $n_{i[-F]}$ √VI

---

[126] I am not claiming that the Spanish speakers I consulted cannot conceive of the thought that there could be

an individual who uses male pronouns that performs the duties of a governess. However, such a thought can

only be expressed via a periphrastic strategy or through the use of a different nominal (e.g., *tutor personal*

'personal tutor').

c. * nP   *Intended*: default

        n √VI
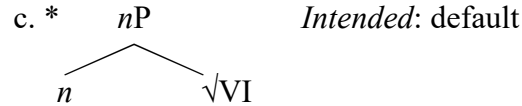
The proposal in (118) accounts for the ill-formed status of (116)c and (117)c: forcing a mismatch would require the √ROOT in the Class VI noun to be licensed by $n_{i[-F]}$ or bare n (both of which would lead to masculine concord on the remnant in NP ellipsis), configurations that violate the proposed licensing conditions for Class VI. The analysis here is identical to the proposal above for Class IV nouns, the sole difference being that there does not seem to exist a grammatically male counterpart for the noun *institutriz* 'governess' in a different class.

4.5.3 Class VII

Class VII is composed of a grammatically masculine noun for which there does not seem to be a grammatically feminine counterpart:[127]

(119)   *Class VII noun*
       músico   'musician'

The denotation of the noun above includes the specification that the individual uses male pronouns. Speakers cannot produce a counterpart of the above and use it for an individual who uses female pronouns. The obvious candidate *música* 'music' is unacceptable or infelicitous:[128]

---

[127] I would consider *gramático* 'grammarian' to be a second member of this class, but none of my consultants use this lexical item. The obvious feminine counterpart *gramática* means 'grammar'. My judgments are that this noun behaves identically to *músico* under ellipsis.

[128] The *Real Academia Española* (RAE) in the 2005 edition of the *Diccionario Panhispánico de Dudas* (accessible at https://www.rae.es/dpd/) claims that the grammatically feminine version of *músico* is, indeed,

(120)   a. *Luisa es músico.
           Luisa is musician
           Intended: 'Luisa is a musician.'

        b. */#Luisa es músic-a
           Luisa   is music-A
           Intended: 'Luisa is a musician.' / *Possible*: 'Luisa is music.'

A [GENDER] mismatch is impossible:

(121)   *Class VII: no mismatch under NP ellipsis*
        a. Los   músicos de mi banda y los   músicos de tu banda no se
           the.M.P musicians of my band and the.M.P musicians of your band   not SE
           llevan.
           get.along
           'My band's musicians and your band's musicians don't get along.'
        b. Los músicos de mi banda y los <músicos> de tu banda no se llevan.'
           'My band's musicians and your band's don't get along.'
        c. *Los músicos de mi banda y las <*ellipsis site*> de tu banda no se llevan.'
           Intended: 'My band's musicians and your band's (female musicians) don't get
                       along.'

(122)   *Class VII: no mismatch under predicate ellipsis*
        a. Tony es músico   y   Jimmy también es músico.
           Tony is musician and Jimmy too   is musician
           'Tony is a musician and Jimmy is a musician too.'
        b. Tony es músico y Jimmy también <es músico>.
           'Tony is a musician and Jimmy is too.'
        c. *Tony es músico y Amanda Miguel también <*ellipsis site*>.
           Intended: 'Tony is a (male) musician, and Amanda Miguel is too.'

Similarly to the analysis of Class VI in the previous subsection, I propose that the noun in

this class contains a √ROOT that can only be licensed by one nominalizer—in this case,

masculine $n_{i[-F]}$:

---

música. Similarly to the case of *institutriz* 'governess', though, my consultants and I would express the

intended meaning via an alternative strategy. For example:

(i)   Laura compone música / se dedica   a   la música/ es instrumentista …
         Laura composes music / SE dedicates PREP the music   is instrumentalist
         'Laura composes music.' / 'Laura's business is music.' / 'Laura is an instrumentalist.'

(123) *Licensing conditions for Class VII nominal*
    a. * nP   *Intended*: feminine

                 $n_{i[+F]}$ √VII

    b. nP   masculine

                 $n_{i[-F]}$ √VII

    c. * nP   *Intended*: default

                 n √VII
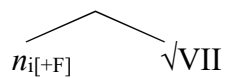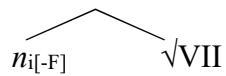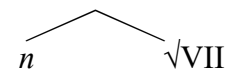
It is conceivable that (some kind of) semantic blocking by the extant expression *música* 'music' could contribute to the existence of the encyclopedic gap I propose here, though I leave exploring this in more detail for the future. Note that Class VII is unlike Class VI in this respect—for Class VI, there is simply no expression at all pronounced *institutor (i.e., 'male governess' or any other meaning).

The ellipsis facts follow straightforwardly—forcing a [GENDER] mismatch here requires there being a Class VII nominal in the ellipsis site that has been licensed by $n_{i[+F]}$. This constitutes a licensing violation that is not repairable by ellipsis.

4.5.4 Class VIII

Class VIII is composed of inanimate nouns. All [GENDER] mismatches are impossible here, a fact that, as we shall see, is straightforwardly derived by the proposal in this dissertation.

Consider now some nouns in this class. Note that some of the noun pairs here are semantically related, while others are not:

(124) *Class VIII noun*
    a. *Nouns with alternating theme vowel*
        manzana/manzano   'apple/apple tree'
        cereza/cerezo   'cherry/cherry tree'
        pesa/peso   'scale/weight'

pasa/paso	'raisin/step'
caso/casa	'case/house'
libro/libra	'book/pound'

    b. *Nouns without an alternating theme vowel*

(el) orden / (la) orden	'order/command'
(el) mar / (la) mar	'sea'[129]

A [GENDER] mismatch is ill-formed with all of these nouns. To illustrate this fact, we can use NP ellipsis and force the mismatch by manipulating determiners in the remnant:

(125)   *Class VIII: no mismatch under NP ellipsis (adapted from Saab 2008)*

      *El   manzano de Pedro es muy viejo, pero la   <manzana> suya
      the.M apple.tree of Pedro is very old   but the.F apple   its.F
      todavía es la   mejor.
      still   is the.F best
      Intended: 'Pedro's apple tree is very old, but its apples are still the best.'

(126)   *Class VIII: no mismatch under NP ellipsis (adapted from Saab 2008)*

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

(127)   *Class VIII: no mismatch under NP ellipsis*

    a. La   cereza que comí ayer   y   las <cerezas> de ese árbol.
      the.F cherry that ate   yesterday and the.F.P cherries of that tree
      son deliciosas
      are delicious.F.P
      'The cherry I ate yesterday and that tree's cherries are delicious.'

---

[129] The semantic difference between the grammatically masculine and feminine versions of *mar* 'sea' is subtle; *la mar* is infused with a poetic quality absent in the masculine version. Juan Uriagereka points out another candidate here: some Spanish speakers allow for a grammatically feminine version of *calor* 'heat'; i.e., there is an alternation *(el) calor / (la) calor*. Neither my consultants nor I accept the grammatically feminine version of this nominal.

b. El   cerezo   es viejo, pero las   cerezas de sus ramas   son
    the.M cherry.tree is old   but the.F.PL cherries of its branches are
    deliciosas.
    delicious.F.P
    'The cherry tree is old, but the cherries from its branches are delicious.'

c. *El cerezo es viejo, pero las <cerezas> de sus ramas son deliciosas.
    Intended: 'The cherry tree is old, but the cherries from its branches are delicious.'

I follow Saab 2004, 2008 and assume that the masculine and feminine versions of these inanimate nouns do not share the same √ROOT. This entails that, from a synchronic perspective, the form-relatedness in these pairs is the result of partial or total homophony. For example, I propose that in the noun pair (*el) orden/(la) orden* 'the order/the command', the √ROOT in the masculine version is licensed only by default n (resulting in masculine concord on determiners and modifiers); conversely, the √ROOT in the feminine version is licensed only by $n_{u[+F]}$—recall our discussion in 4.3.2 on the licensing heads that combine with inanimates:[130]

(128)   *Sole licensing condition for el orden 'the order'*

        *n*P   default

    n   √66

(129)   *Sole licensing condition for la orden 'the command'*

        *n*P   uninterpretable feminine

    $n_{u[+F]}$ √888

---
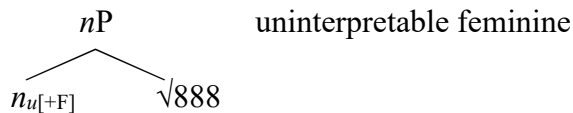
[130] Note that Saab (2004): 40 proposes that the masculine version of *orden* 'order' combines with a head bearing masculine features.

A [GENDER] mismatch under ellipsis as in (126) is ill-formed, then, because this configuration violates the second statement in the identity condition, which requires absolute √ROOT identity.[131]

4.5.5 Class IX

Nouns in Class IX control feminine concord but denote humans of any gender:

(130)   *Class IX nouns*
    víctima   'victim'
    persona   'person'
    criatura   'child'

Observe that nouns in this class can be used predicatively with individuals understood to use any set of pronouns:

(131)   *Class IX nouns*
    a. Lupita / Chepe es persona de confianza.
       Lupita   Chepe is person   of trust
       'Lupita / Chepe is a trustworthy person.'

    b. Lupita / Chepe fue víctima de un asalto.
       Lupita   Chepe was victim   of a   robbery
       'Lupita / Chepe was victim of a robbery.'

---

[131] One might be worried by my proposal that noun pairs that are semantically related like *manzana/manzano* 'apple/apple tree' and, perhaps more worryingly, *(el) mar/ (la) mar* 'sea' do not share the same √ROOT (see Saab 2010a, fn. 6). Omer Preminger suggests an alternative where pairs like these do, in fact, share the same √ROOT. Take *manzana/manzano*. The feminine would be licensed by $n_{u[+F]}$, while the masculine would be licensed by masculine $n_{i[-F]}$. However, this proposal would analyze *manzano* as an idiom, where the Encyclopedic insertion for the √ROOT in the context of $n_{i[-F]}$ is the interpretation 'apple tree', not any denotation of, say, a sexed apple. Under this account, the ill-formed mismatches would result from *n* clashes, violating the identity condition proposed here. I leave an assessment of this alternative for future work, since it raises the broader issue of whether we need uninterpretable [GENDER] features at all; i.e., all inanimates could be licensed by *interpretable* [GENDER] features and be idioms.

Forcing a [GENDER] mismatch can only be tested here via NP ellipsis, then, and as shown below, this mismatch is ill-formed:

(132)   *Class IX: no mismatch under NP ellipsis*
      a. La   víctima del   asalto y   la víctima del   fraude declararán hoy.
          the.F victim of.the robbery and the victim of.the fraud declare.FUT today
          'The victim of the robbery and the victim of the fraud will declare today.'
      b. La víctima del asalto y la <víctima> del fraude declaran hoy.
      c. * La víctima del asalto y el *<ellipsis site>* del fraude declaran hoy.

I propose that nouns in this class contain a √ROOT that is licensed exclusively by $n_{u[+F]}$:

(133)   *Sole licensing condition for Class IX nouns*

$$nP \qquad\qquad \text{uninterpretable feminine}$$
$$n_{u[+F]} \qquad \surd IX$$
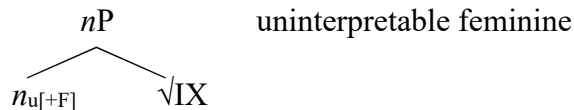
In a mismatch example like (132)c, then, the ellipsis site contains a Class IX √ROOT licensed by either a masculine $n_{i[-F]}$ or plain *n*. However, this nominal in the ellipsis site violates the licensing condition for a Class IX √ROOT and this violation cannot be repaired by ellipsis—the ill-formed status of the mismatch is thus derived.

4.5.6 Class X and Class XI

In this subsection, I discuss two classes of nouns whose members all denote animals. Mismatches are ill-formed across the board here as well.

    Let us start with Class X, which is composed of grammatically feminine nouns that denote animals. Though these nouns are grammatically feminine, their denotation does not include a sex specification. Instead, the way to express a specific sex in these cases is through the use of the modifier *macho* 'male' or *hembra* 'female'. A sample of nouns in this class is given below:

(134) *Class X nouns*

serpiente	'serpent'	serpiente macho/hembra	'male/female serpent'
culebra	'snake'	culebra macho/hembra	'male/female snake'
ballena	'whale'	ballena macho/hembra	'male/female whale'
danta	'tapir'	danta macho/hembra	'male/female tapir'

[GENDER] mismatches under ellipsis are ill-formed with Class X nouns:

(135) *Class X nouns: no mismatch under NP ellipsis*

a. Se escaparon la   culebra de Elena y   la   culebra de Olga.
   SE escaped   the.F snake   of Elena and the.F snake   of Olga
   'Elena's snake and Olga's snake escaped.'

b. Se escaparon la culebra de Elena y la <culebra> de Olga.
   'Elena's snake and Olga's escaped.'

c. * Se escaparon la culebra de Elena y el <*ellipsis site*> de Olga.
   Intended: 'Elena's snake and Olga's (male snake) escaped.'

My analysis here is identical to the one proposed for Class IX in the previous subsection.

Namely, Class X nouns contain a √ROOT that is licensed by $n_{u[+F]}$ only:

(136) *Sole licensing condition for Class X nouns*

nP   uninterpretable feminine

$n_{u[+F]}$ √X
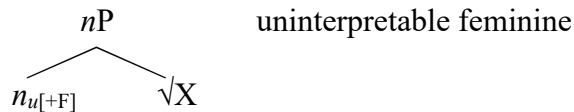
The ill-formed status of the ellipsis data in (135)c follows from the impossibility of Class X nouns being composed of a √ROOT that is licensed by $n_{i[-F]}$ or plain n—recall, once more, that I am proposing that a violation of licensing conditions on √ROOTs cannot be repaired by ellipsis.

It is worth pausing at this juncture to reflect on the relationship between morphosyntactic structure and encyclopedic content by highlighting a contrast between Class IX (e.g., *la víctima* 'the victim') and Class X nouns. In brief, even though the members of each of these classes differ significantly in their meaning—the former all denote humans, the latter, animals—their morphosyntax is identical (i.e., they are

composed of √ROOTs that are licensed solely by $n_{u[+F]}$; see (133) and (136)). What the present proposal entails, then, is that many classes of expressions can arise from a combination of morphosyntactic structure and encyclopedic content, even though the system contains only four variants of categorizing *n*.

Let us move on to Class XI. The nouns in this class also denote animals but are grammatically masculine. In a manner parallel to nouns in Class X, the denotation of these nominals does not include a sex specification. The way to specify this is through the same modifiers that were discussed in relation to Class X:

(137) *Class XI nouns*

delfín	'dolphin'	delfín macho/hembra	'male/female dolphin'
cuervo	'crow'	cuervo macho/hembra	'male/female crow'
tapir	'tapir'	tapir macho/hembra	'male/female tapir'
tepezcuintle	'lowland paca'	tepezcuintle macho/hembra	'male/female lowland paca'

Much like our previous discussion of Class X nouns, a [GENDER] mismatch is impossible with Class XI nouns:

(138) *Class XI nouns: no mismatch under NP ellipsis*

a. Están en peligro de extinción el   delfín de agua dulce y   el   delfín
   are   in danger of extinction the.M dolphin of water sweet and the.M dolphin
   de agua salada?
   of water salty
   'Are the freshwater dolphin and the saltwater dolphin endangered?'

b. Están en peligro de extinción el delfín de agua dulce y el <delfín> de agua salada?

c. *Están en peligro de extinción el delfín de agua dulce y la <*ellipsis site*> de agua salada?
   Intended: 'Are the freshwater dolphin and the seawater (female) dolphin endangered?'

The way to capture the above facts is that Class XI nouns contain a √ROOT that is licensed by default *n* only:

(139) *Sole licensing condition for Class XI nouns*

nP   default

n √XI
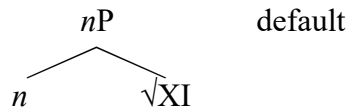
The ill-formed status of the ellipsis examples is derived in a manner parallel to the proposal for Class X above—a mismatch here requires an ellipsis site containing a Class XI noun whose √ROOT is licensed by $n_{i[+F]}$ or $n_{u[+F]}$, in violation of the proposed licensing conditions for Class XI nouns. Ellipsis cannot repair these violations, so a [GENDER] mismatch is ill-formed.

4.5.7 Final comments

My proposal that the identity condition on ellipsis includes a statement requiring featural non-distinctness raised the following question: if, as assumed in this chapter, there is a nominalizing n in Spanish that bears no [GENDER] features (see 4.3.2), should we not expect feminine–masculine mismatches to be broadly available, given that a plain n in the ellipsis site would lead to a configuration that satisfies the identity condition? Empirically, I showed on the basis of a wide array of nouns with different semantic and grammatical characteristics that we do not, in fact, find that such mismatches are always possible. My analysis of this was as follows: some configurations may comply with featural non-distinctness and satisfy the identity condition, but they incur in licensing violations between the nominalizing n and the √ROOT, creating problems in the Encyclopedic component that cannot be repaired by ellipsis. I hope that the discussion here will entice more work assessing whether the unavailability of certain conceivable mismatches can be analyzed in the same manner and provide support for the idea that not all gaps are created equal—some are morphophonological and repaired by ellipsis, while others are not.

4.6 Previous analyses

In this section, I briefly address some of the analyses that have been offered in the past to account for different subsets of the empirical picture that I have dealt with in this chapter. The discussion here should be understood in the context of this chapter and the previous chapters in this dissertation; specifically, in the context of our previous conclusions that an identity condition predicated on featural non-distinctness and √ROOT identity delivers the correct empirical coverage independently of the phenomenon of [GENDER] mismatches. In a way, then, it would be sufficient to criticize the approaches simply because they rely on stricter identity conditions (Saab 2010a; Donatelli 2019). Nevertheless, I will summarize the key elements of the proposals and point out some of the challenges that they face.

The table below summarizes the analyses that I will touch on. Naturally, the table is not exhaustive, but it represents the kind of approaches that have been put forth in the literature so far. Note that the terminology used by Bobaljik & Zocca 2011 and Donatelli 2019 is not the same as the terminology used in this dissertation—in comparing the approaches, this should be kept in mind. I will continue to refer to the classes in the same way as I have in this chapter—i.e., I consider Class I the non-alternating *tío/tía* class; Class II the alternating *abogado/abogada* class; and Class III the asymmetrical *actor/actriz* class:[132]

---

[132] An assessment of Sudo & Spathas 2016 and 2020a will have to be left for the future, since I became aware of the second paper in the latter stages of the writing of this work. A particular difficulty arises in assessing this body of work, because the latter article presents a modified empirical picture from the former, on the basis of a broader sampling of Greek speakers. In fact, Sudo & Spathas (2020)a question the validity of positing the existence of a Class III under ellipsis, given that judgements are particularly variable and subtle

(140) *A snapshot of previous analyses*

	Ellipsis Type	*Identity Condition*	*Description of analysis*
Saab 2010a	NP ellipsis	Strict syntactic identity	NP ellipsis is *n*P ellipsis; [GENDER] cannot mismatch inside ellipsis site
Bobaljik & Zocca 2011	NP ellipsis and predicate ellipsis	Strict identity	Class I nouns are both specified for [GENDER]; Class II nouns are not and gender marking is inflectional; the extra morpheme on Class III feminine nouns is specified for [GENDER].
Merchant 2014	NP ellipsis and predicate ellipsis	Semantic; e-GIVENness (only evaluated for cases of PF deletion; null proforms are not regulated by the condition)	Missing nouns involve ellipsis *or* a null proform; [GENDER] mismatched examples always involve proform e_N; the tripartite asymmetry arises from the possibility/impossibility (dependent on noun class membership) of the e_N locating an antecedent that conflicts or does not conflict with the subject of the target clause.
Donatelli 2019	Predicate ellipsis	Proper subset condition (Murphy 2016)	Gender on Class II nouns is interpreted as presuppositional; gender on Class I nouns is assertional and in the root; gender on feminine Class III nouns is assertional and on the derivational affix.

Let us start with Saab 2010a, which discusses NP ellipsis in Spanish and focuses primarily on the asymmetric behavior of [GENDER] in opposition to [NUMBER] (see 4.2 above). Specifically, Saab analyzes a set of data showing that mismatches in the former feature are ill-formed, whereas the latter feature is allowed to mismatch freely. Based on this, Saab

---

(see 4.2.1). I do note that the authors show that independent empirical domains demonstrate the existence of three distinct classes of nouns in Greek, independently of ellipsis, and they provide a battery of diagnostics that would be useful to replicate in other languages. Perhaps these independent signals of noun class membership can provide a clue as to how these classes arise under ellipsis in other languages, shedding some light on the tricky question of how learners come to acquire the relevant distinctions (see 4.9).

proposes that NP ellipsis targets *n*P to the exclusion of NumP (see 4.3.1). Furthermore, he proposes a two-pronged identity condition (see Saab 2004, 2008). Note, as I have mentioned before, that the partitioning of the final identity condition in this dissertation into a statement pertaining to features on the one hand, versus √ROOTs on the other, is directly inspired by Saab's proposal:

(141)   *Saab 2010a's identity condition*
        A constituent C can be elided if there is a constituent C′ identical to C in the syntax. Identity:
        (A) An abstract morpheme α is identical to an abstract morpheme β iff α and β match all its semantic and syntactic features.
        (B) A Root A is identical to a Root B iff A and B have the same label.

Returning to NP ellipsis, the asymmetry between [NUMBER] and [GENDER] that Saab discusses is derived under his approach because *n* is assumed to be the locus of [GENDER] (just like I assume; see 4.3.2); since *n* is inside the ellipsis site in NP ellipsis, any mismatch in *n* flavors leads to a violation of featural identity. Conversely, since [NUMBER] is outside the ellipsis site, it can mismatch freely.

The main issue with the proposal in Saab 2010a is its limited empirical coverage. While it derives the Spanish data reported therein—which show that [GENDER] mismatches are ill-formed across the board under NP ellipsis[133] —we know that this is not the case cross-linguistically. The limited applicability of the analysis is a problem for any unified account of the identity condition. Particularly challenging for Saab's account is the tripartite asymmetry that arises in several grammars and is contingent on noun class membership (Merchant 2014, Alexiadou 2017, Sudo & Spathas 2016, 2020a,b, Bobaljik & Zocca 2011,

---

[133] See section 4.7 below as well for some comments on the difficulty of assessing whether there is no asymmetry *whatsoever* between different nominals for the grammars analyzed in Saab 2004, 2008, 2010a.

Polinsky 2020; see section 4.2 above). Whereas the identity condition proposed here was able to account for grammars of this kind, it is unclear how the strict condition assumed in Saab 2010a could deliver the right result. Additionally, the requirement that features in the ellipsis site be identical to an antecedent in order to be elided is too strong and undergenerates—recall my analysis in chapter 2 of AF-active mismatches in Kaqchikel, as well as several of the mismatch types discussed in chapter 3 (polarity, modality, etc.).

Moving on to Bobaljik & Zocca 2011, these authors use ellipsis to probe the nature of grammatical gender in Brazilian Portuguese. Focusing on predicate ellipsis only, they show that the language exhibits the now-familiar tripartite asymmetry in [GENDER] mismatch. The ingredients of their analysis are as follows: first, they capitalize on a morphological aspect of the class of nouns that allows mismatches symmetrically (our Class II)—all the nouns in this class display an alternation in theme vowel. Namely, masculine nouns end in -*o*, whereas feminine nouns end in -*a* (in a way that is parallel to some Spanish nouns). Bobaljik and Zocca note that these -*o* and -*a* endings also surface on adjectives exhibiting concord. They thus propose that the -*o* and -*a* endings on the noun pairs that allow mismatches signal that gender is not lexically encoded on these nouns. Rather, the authors propose that gender is purely inflectional here and arises via agreement. Second, the authors assume that inflectional mismatches are allowed under ellipsis. Third, they assume that the identity condition on ellipsis is strict and calculated at LF. In a nutshell, there is no mismatch in examples like 'Pedro is a doctor and Marta is too' (9)a and 'Marta is a doctor and Pedro is too' (9)b:

(142)   *Symmetrical mismatch: Bobaljik & Zocca 2011: 159*

    a. LF: O Pedro é médic-   e a Marta também é <médic->.
       Morph: médic-o
    b. LF: A Marta é médic-   e o Pedro também é <médic->.
       Morph: médic-a

Moving on to the non-alternating class (our Class I)—which in Brazilian Portuguese includes kinship nouns and nobility terms—the authors blame the ill-formed status of the mismatched examples on the semantics of the nouns. In brief, both the masculine and feminine versions of these nouns have a semantic gender specification that clashes under ellipsis.[134] Finally, in the case of the asymmetrical licensing class (our Class III), the authors propose that the masculine carries no lexical specification for gender, but the feminine does (as signaled by the additional morphology). This gives rise to the asymmetrical pattern.

Bobaljik and Zocca's proposal faces several challenges. As the authors acknowledge, several nouns in the non-alternating class (our Class I) also display a theme vowel (e.g., the kinship terms for '*uncle/aunt*', '*brother/sister*', etc.). They must assume that for those nouns, the vowel is *not* an agreement marker. Second, and most importantly, their account cannot explain NP ellipsis data for Brazilian Portuguese. Recall that Nunes & Zocca 2009 showed that NP ellipsis also exhibits [GENDER] mismatch asymmetries depending on noun class membership. I repeat the relevant examples below:

---

[134] The authors claim that it is immaterial whether this lexical specification is part of the lexical semantics of the root, or whether a feature encodes it (see their footnote 12). If the analysis presented in this chapter is on the right track, I have shown that this kind of nuance does matter, since I have capitalized on the locus of [GENDER] in the structure—on the nominalizer *n*—as well as the interaction of this head with specific √ROOTs (our licensing conditions; see 4.3.2).

(143) *[GENDER] mismatch symmetrically ill-formed; Brazilian Portuguese*

    *O   João visitou o   tio   dele   e   o   Pedro prometeu visitar a   dele.
    the João visited the uncle of.his and the Pedro promised visit   the.F.SG of.his
    Intended: 'João visited his uncle and Pedro promised to visit his aunt.'

(144) *[GENDER] mismatch symmetrically well-formed; Brazilian Portuguese*

    O   João visitou o   médico   dele   e   o   Pedro visitou a   dele.
    the João visited the doctor.M.SG of.his and the Pedro visited the.F.SG of.his
    'João visited his doctor and Pedro visited his (female) doctor.'

The aspect of the examples above which proves difficult to reconcile with Bobaljik & Zocca 2011's account is that the remnant determiner in the well-formed example in (144) displays a reflex of a feminine feature, a feature which must be borne by the elided nominal in the ellipsis site. Recall that Bobaljik & Zocca assume that nouns like 'doctor' do *not* have a gender feature when used predicatively, and this is why they do not create an issue for the identity condition under predicate ellipsis (see (142)). In the case above of *argument ellipsis*, though, we observe the same pattern as with predicate ellipsis—crucially, there is a [GENDER] feature in the ellipsis site triggering concord on the remnant determiner.

Bobaljik & Zocca 2011's footnote 22 acknowledge this problem. There, the authors address a reviewer's question about argument ellipsis. Even though they do not describe what the pattern actually *is* with NP ellipsis, they suggest that whatever needs to be said about adjectives modifying zero nominals in argument position would explain the facts— e.g., in the Spanish *Los ricos también lloran*; 'The rich also cry.' It is difficult to evaluate this suggestion, so I maintain that Bobaljik & Zocca's analysis faces serious challenges as an account of the empirical picture presented in this chapter.

Moving on to Merchant 2014, the author describes an apparent asymmetry in Greek, wherein predicate ellipsis shows our familiar tripartite asymmetry and NP ellipsis bans [GENDER] mismatches altogether. Merchant argues that the syntax of all nominals is

identical, with the difference among classes of nouns lying in the gender information that they are semantically specified for: each member of a Class I pair is specified for gender; Class II nouns are not; and finally, the feminine in a Class III pair is specified, while the masculine is not:

(145)   *Lexical specifications (adapted from Merchant 2014 (39-(41))*

a. [[adherfos]] = $\lambda X_e : X$ is male [sibling(X)]
b. [[adherfi]] = $\lambda X_e : X$ is female [sibling(X)]   *Class I*
c. [[jatros]] = λX_e [doctor (X)]   *Class II*
d. [[dhaskalos]] = λX_e [teacher(X)]
e. [[dhaskalos]] = $\lambda X_e : X$ is female[teacher(X)]   *Class III*

Merchant proposes that there are two paths to silent nouns. One involves true ellipsis and requires strict semantic identity (e-GIVENness; Merchant 2001). When there is a [GENDER] mismatch, this strategy is not available, since the identity condition is violated (see above on the denotations of different nouns). As a result, Merchant proposes that all cases where there is a *well-formed* [GENDER] mismatch do not involve true ellipsis, but a null pro-form e$_N$ that is analogous to English *one.*

Why are all NP ellipsis examples ill-formed, then, since a pro-form e$_N$ could be used? In a nutshell, Merchant argues that NP ellipsis with a [GENDER] mismatch is ill-formed across the board because the pro-form e$_N$ strategy leads to a morphosyntactic problem: determiners or modifiers in the remnant of a target clause cannot have their φ-features valued by e$_N$, so the derivation crashes.[135]

---

[135] Why this doesn't lead to the insertion of default nominal concord, instead of a derivational crash, is not addressed by the author.

This leaves predicate ellipsis, where some mismatches are well-formed and others are ill-formed. Merchant attributes the ill-formed examples to a semantic anomaly arising from the resolution of e_N to its antecedent. Consider a Class I mismatch, which is symmetrically ill-formed:

(146) *Class I mismatch: masculine – feminine (Merchant 2014: example (65)a)*
      *O   Petros ine kalos   adherfos, ala i   Maria ine mia kakia $<e_N>$.
      the Petros is   good.M brother   but the Maria is a.F bad.F
      Intended: 'Petros is a good brother, but Maria is a bad sister.'

The problem, according to Merchant, lies in the following: the interpretation of e_N in the "ellipsis site" is resolved to *adherfos* 'brother' in the antecedent, which has a specification denoting someone who uses male pronouns (adapting Merchant 2014's terminology). This creates a semantic anomaly in the second clause, since the subject is *Maria*. The same analysis rules out feminine-masculine mismatches with Class I and Class III. On the other hand, well-formed mismatches are ruled in because the interpretation of e_N is resolved to a nominal in the antecedent that does not carry a lexical specification. As a result, no clash results with the subject of the second clause and the examples are well-formed.

I noted previously that Sudos & Spathas 2016 and Alexiadou 2017 consider that the putative asymmetry between NP ellipsis and predicate ellipsis in Greek is not empirically supported, since all of Merchant's NP ellipsis examples exhibited independent confounds that rendered them ill-formed. Nevertheless, we can assess the flavor of the analysis independently, and point out several challenges. First, it cannot explain grammars where the tripartite asymmetry is observed in both NP ellipsis and predicate ellipsis—i.e., the Spanish grammar described in this chapter, as well as the grammars reported by other scholars working on Brazilian Portuguese and Greek. Furthermore, Merchant 2014's contention that putative ellipses with [GENDER] mismatches involve a null pro-form is

difficult to reconcile with the possibility of sub-extraction in Spanish mismatched examples (see 4.3.1); furthermore, Sudo & Spathas 2016, 2020a show that NP ellipsis examples in Greek with [GENDER] mismatches and sub-extraction are deviant even in their non-elliptical counterparts, suggesting that sub-extraction cannot be used as a diagnostic for pro-form versus true ellipsis in these Greek cases, in the first place. Lastly, Merchant 2014's proposal raises the following question: is the pro-form strategy for silence always available if the true ellipsis strategy fails due to a violation of identity, or does this have something to do with grammatical gender specifically? In other words, do identity violations related to other features—e.g., voice mismatches under sluicing that involve a clash of Voice[ACT] vs. Voice[PASS]—lead to the possibility of a second, pro-form strategy for silence that must also be ruled out?

Finally, Donatelli 2019 discusses predicate ellipsis in Spanish and adopts the proposal that the identity condition on ellipsis involves a subset relation: ellipsis proceeds if the ellipsis site is a featural subset of the antecedent (Murphy 2016, Saab 2015):[136]

(147)   *Subset condition on ellipsis (Donatelli 2019)*
        The morphosyntactic features in the ellipsis site must be a proper subset of those in the antecedent ($F_E \subset F_A$)

Donatelli further assumes that Class II feminine nominals bear [GENDER] features, but gender is interpreted as presuppositional, while it is interpreted as assertional for Class I and feminine Class III (see Donatelli 2019: 170; recall the difference in terminology).

---

[136] This condition is attributed in Donatelli 2019: 177 to Saab 2015. However, it seems that only Murphy 2016 proposes such a condition. As far as I can tell, Saab 2015 does not state that a proper subset relation *must* be satisfied under ellipsis. Instead, Saab states that a proper subset relation is *enough* to satisfy the identity condition.

Finally, Donatelli assumes that features that result in an interpretation of [GENDER] as presuppositional can be ignored under ellipsis.

Several issues arise with Donatelli's account. First, it is unclear why the ultimate interpretation of [GENDER] features as either presuppositional or assertional would matter under Donatelli's conception of the identity condition, which appears to be syntactic in nature. Under the proposal advocated in this dissertation, no such stipulation needs to be made, since the condition on features is satisfied by non-distinctness. Second, we have seen that a proper subset condition like the one adopted by Donatelli undergenerates, given (among other cases) the possibility of polarity mismatches where there is negation in the ellipsis site and no negation in the antecedent (chapter 2), as well as the AF-Active mismatches in Kaqchikel. The latter examples involve a head bearing a feature (Voice$_{[ACT]}$) in the ellipsis site and no such head (bearing any feature) in the antecedent. In fact, Donatelli's condition seems to ban ellipsis whenever the features in the ellipsis site are *identical* to the antecedent—note that the condition requires a *proper* subset relation (I discuss this more in section 5.7 of this dissertation). This is undesirable, since strict identity clearly satisfies the identity condition (e.g., '*She lip-synced to three **songs**, while I lip-synced to four <**songs**>*' is well-formed).

To summarize, I went over four prior analyses for the phenomenon discussed in this chapter—Saab 2010a, Bobaljik & Zocca 2011, Merchant 2014, and Donatelli 2019. I presented a snapshot of the proposals and pointed out the main issues faced by each given the overall empirical picture attested cross-linguistically regarding [GENDER] under ellipsis, as well as the conclusions drawn in the previous chapters of the present work. I was unable

to address every single facet of these proposals, since it would take us too far afield, but I hope to have laid out the issues in need of consideration.

4.7 Microvariation

In investigating [GENDER] mismatches under ellipsis, one observation recurs time and again throughout the literature. Authors remark, usually as a brief aside, that there appears to be significant microvariation across multiple axes. Consider the quotation below from Merchant 2014, which lays out some of the core issues surrounding microvariation that should be addressed by any analysis of the phenomenon:

(148) *Merchant 2014 (footnote 6) on 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 nonalternating class). A fuller exploration of the variation in this domain is needed."

This quotation is revealing in two ways. First, when placed in the context of our current investigation, it reveals that there is *cross-linguistic* variation in the classification of specific noun pairs and their concomitant behavior under ellipsis—for instance, Merchant's main consultant classifies the kinship terms 'uncle/aunt' in the one-way alternating class (i.e., our Class III), whereas for my Spanish consultants, this pair is Class I (see Bobaljik & Zocca 2011, which also reports 'uncle/aunt' as disallowing mismatches symmetrically for Brazilian Portuguese). Merchant's observation on the variable classification of the noun pair 'uncle/aunt' already casts doubt on any proposal that would seek to blame the *semantics* of kinship nouns as being responsible for their behavior (see 4.6 above on Bobaljik & Zocca 2011). The second way in which this quotation speaks to

314

the issue of microvariation is by pointing out that *intra-linguistically*, different speakers may assign noun pairs to different classes—i.e., while the author's main consultant places 'uncle/aunt' in the one-way alternating class, another of the author's informants considered the pair to be in the nonalternating class (i.e., our Class I—in parallel to the Spanish data I presented and Brazilian Portuguese).

This section addresses microvariation, discussing three issues and setting the stage for a more in-depth investigation in the future. My purpose here is not to provide an account for every single point of variation that has been reported. Rather, I intend to show that (i) the flexibility offered by a decompositional approach to nominals, (ii) repair-by-ellipsis, and (iii) the identity condition proposed here, when taken together, furnish the necessary ingredients for an account that would capture the range of grammars that seem to be attested—though questions remain, of course (see 4.7.2 in particular).

In 4.7.1, I first discuss the issue of noun classification cross- and intra-linguistically, putting together in one place some of the variability that has been reported. I point out how we can derive the patterns we observe by keeping the identity condition *constant* and attributing the variation instead to speakers' different representations for specific lexical items. I also comment briefly on some of the (open) questions regarding acquisition that stem from the attested microvariation. In 4.7.2, I discuss that, for some speakers, there might be a contrast between elliptical *constructions* in relation to [GENDER] mismatches: i.e., NP ellipsis disallowing mismatches (perhaps altogether) and predicate ellipsis exhibiting our familiar tripartite asymmetry. I show first the difficulty in establishing which speakers might possess such a grammar, given that authors focus almost exclusively on predicate ellipsis and do not provide full, controlled data samples for NP ellipsis alongside

their description of predicate ellipsis patterns. Nevertheless, evidence from Saab 2004, 2008 and Masullo & Depiante 2004a,b suggests that a grammar of this type is possible. I leave as an open question how such a grammar could be accounted for, but suggest a solution.[137]

4.7.1 Different speakers, different classes of nouns

The issue of variable assignment into classes of nouns can be illustrated by contrasting the Spanish data I presented in this chapter with prior, and occasionally conflicting reports in the literature. For example, while Bobaljik & Zocca 2011 focuses on Brazilian Portuguese, brief remarks on Spanish and a list of nouns corresponding to two different classes of nouns are provided.[138] The authors claim that none of their Spanish consultants seem to have a class where [GENDER] mismatches freely under ellipsis; i.e., in our terms, none of their speakers has a Class II. This contrasts with the report here, as well as with Donatelli 2019's, which provides the following data (I return to the (?) in (149)b later in this section):

(149) *Donatelli 2019: 127—alternating class (our Class II)*
    a. Pablo es médico y   Marta también.
       Pablo is doctor.M and Marta too
       'Pablo us a doctor and Marta is too.'
    b.(?)Marta es médica y   Pablo también.
        Marta is doctor.F and Pablo too
       'Marta is a doctor and Pablo is too.'

Bobaljik & Zocca do not provide any Spanish judgements (except for one ill-formed NP ellipsis example with *tío/tía*—Class I for us—attributed to Kornfeld & Saab 2004; see their

---

[137] For a discussion of the methodological issues that arise in eliciting judgements on grammatical gender alternations, see Donatelli 2019: section 6.5.

[138] The authors, however, do not specify the dialect(s) spoken by their consultants.

fn. 10). In an appendix, however, they provide a sample of Spanish noun pairs and their classification. There are only two pairs that denote professions in their list—the semantic trait that seems to unify our Class II—but these are classified in the asymmetric licensing class (our Class III):

(150) *Spanish nouns denoting professions—Bobaljik & Zocca 2011*
       director / director**a**   'male director/female director'
       profesor / profesor**a**   'male professor/female professor'

Note that for these noun pairs, the masculine does not bear a theme vowel, while the feminine does. Interestingly, the authors do not provide any noun pairs where both masculine and feminine bear a theme vowel (e.g., our by-now familiar *abogado/abogada* 'lawyer' or *médico/médica* 'doctor').

I propose that the analysis in this dissertation can account for the empirical picture sketched by Bobaljik & Zocca for their Spanish consultants. Suppose that for these speakers, the masculine in a pair like *director/directora* is licensed by plain *n*. Suppose, furthermore, that the *-a* vowel on the feminine noun is represented by these speakers as a √ROOT akin to *-iz* in *actriz* (crucially, Bobaljik & Zocca give *actor/actriz* 'actor/actress' as a member of the same class for their Spanish speakers). The analysis I put forward in 4.4 can thus derive the facts: in the well-formed masculine-feminine mismatch, there is repair-by-ellipsis, whereas the inverse is ill-formed due to a violation of strict √ROOT identity. The variation between the Spanish speakers consulted by Bobaljik & Zocca and my own (as well as Donatelli's) boils down to a different structural representation of the specific

nominals under consideration. In brief, the variation does not need to be encoded via any difference in the identity condition on ellipsis.[139]

With this discussion in mind, we can now return to example (149). In relation to this manipulation, Donatelli provides the following comment on the (?) notation given for the feminine-masculine mismatch:

(151) *Donatelli 2019 on alternating class—our Class II[140]*
"(1b) is acceptable for the majority of speakers, though some slightly disprefer it to (1a). The variable acceptability of (1b) tends to be in stark contrast to the unacceptability of (2b) [***author comment***: a feminine-masculine mismatch with *actriz-actor* (our Class III)]. However, it seems some speakers treat example (1) exactly as (2), allowing for the (a) constructions but not the (b) constructions…"

Similar remarks are made by Saab 2004 on the basis of parallel data. First, I present the data Saab discusses (152)a-b, followed by the relevant comment (translations mine from Spanish):

(152) *Saab 2004: 51, footnote 20*
a. Juan es maestro   y   María también
   Juan is teacher.M and María also
   'Juan is a teacher and María is too.'

---

[139] Bobaljik & Zocca report that animals fall in the asymmetric class as well for their speakers, providing the pairs *gato/gata* 'cat' and *oso/osa* 'bear'. In this case, I would assume that for their speakers (i) the masculine version of these animal denoting nouns is licensed by plain *n*, and (ii) the -o vowel on the masculine is a theme vowel, whereas the -a on the feminine is also represented as a √ROOT. The facts then follow from the identity condition in the same way as with the human denoting nouns in (150)—see below in the main text on Donatelli 2019's comment on the examples in (149).

[140] The issue is brought up again in Donatelli 2019: 144, where the author notes that speakers who dislike the feminine – masculine mismatch with these noun pairs are from Spain and Chile. Nevertheless, a remark is also made that the contrast with asymmetric 'actor/actress' pairs (our Class III) is robust for all speakers.

b. # María es maestra y   Juan también.
   María is teacher.F and Juan too
   'María is a teacher and Juan is too.'

c. "When the first coordinate is feminine, some speakers find the sentence anomalous… To our judgment, the sentence is perfect, but we do not have an explanation as to why for certain speakers, it is not."

It is possible that the speakers that Donatelli and Saab allude to have a similar grammar as the speakers consulted in Bobaljik & Zocca 2011—in brief, all these speakers classify noun pairs denoting professions as our Class III. My analysis would be straightforward and provides, perhaps, the explanation that is sought by Saab—for these speakers, (i) the masculine noun is licensed by plain *n*, (ii) the *-o* is a theme vowel, and (iii) the *-a* exhibited by the feminine is represented as a √ROOT.

In the context of my proposal for this pocket of microvariation, a brief aside into Russian illustrates the same sort of phenomenon and the plausibility of extending the account beyond Spanish. First, consider that Bobaljik & Zocca 2011 provides the following Russian judgement. In this case, the nouns seem to behave like our Class III:

(153) *Russian—data adapted from Bobaljik & Zocca 2011*
   a. Ivan moskovič   i   Marina tože <moskovič-ka>.
      Ivan Muscovite.M and Marina too   Muscovite-F
      'Ivan is a Muscovite and Marina is too.'
   b. # Marina moskovička I Ivan tože <moskovič>.
      Intended: 'Marina is a Muscovite and Ivan too.'

The data above could be captured by an account parallel to the one I proposed earlier for Spanish Class III nouns. Namely, the Russian speakers consulted by Bobaljik & Zocca consider the feminine affix *-ka* a √ROOT, yielding the asymmetrical pattern that has become familiar by now. Interestingly, we observe microvariation with noun pairs like those in (153); some speakers treat these examples on par with our Class II examples and a mismatch is symmetrically well-formed (Maria Polinsky p.c.). For those speakers, I

propose that the *-ka* morpheme on the feminine is not represented as a √ROOT, but is akin to the theme vowels on Spanish nouns like *abogado/abogada* 'lawyer'.

Before ending, I will briefly address a particular group of nouns that share a semantic commonality and which most authors place in the non-alternating class (our Class I; see Bobaljik & Zocca 2011, Saab 2008, Donatelli 2019)—these are nobility terms:

(154)   *Nobility terms*
        príncipe/princesa   'prince/princess'
        rey/reina   'king/queen'
        duque/duquesa   'duke/duchess'
        marqués/marquesa   'marquis/marquise'
        conde/condesa   'count/countess'
        barón/baronesa   'baron/baroness'

In the case of *príncipe/princesa*, it seems that we are dealing with a shared √ROOT and both members of the pair bear an additional exponent; for all other cases, only the feminine appears to bear an additional exponent. For speakers who consider all of these as Class I, the analysis is straightforward—under my account, these speakers analyze the feminine as licensed by $n_{\text{i}[+\text{F}]}$ and the masculine as licensed by $n_{\text{i}[-\text{F}]}$. An ellipsis example leads to a clash of features and is ruled out by the identity condition proposed in this dissertation.

However, one other consultant and I judge there to be a contrast between *príncipe/princesa* and *rey/reina* on the one hand, and all other nobility pairs on the other hand—the latter being those where the feminine bears a *-sa* morpheme. Mismatches are symmetrically ill-formed with 'prince/princess' and 'king/queen', whereas we would classify pairs like *duque/duquesa* 'duke/duchess' as Class III, since there is a contrast in acceptability between a case of mismatch where the masculine is in the antecedent, versus in the ellipsis site (see footnote 102). Tellingly, perhaps, Donatelli 2019: 196 shows judgments for [GENDER] mismatches with *marqués/marquesa* and *conde/condesa* where

the masculine-feminine mismatch is judged ?*, whereas the feminine-masculine is judged *. On the other hand, a mismatch with *príncipe/princesa* is judged * symmetrically. The analysis proposed in this chapter can provide the right kind of flexibility to accommodate this sort of variation—for some speakers, the masculine in these noun pairs is licensed by plain *n* and the *-sa* morpheme on the feminine is analyzed as a √ROOT, giving rise to the asymmetrical Class III pattern (see the noun pair *alcalde/alcaldesa* 'mayor' in 4.4.2).

In sum, the microvariation we observe can be attributed to the existence of different lexical representations between speakers for specific noun pairs. Naturally, though, several questions related to acquisition arise as a result of this account (see a discussion of similar concerns in Donatelli 2019). First, it seems reasonable to assume that learners of Spanish, Greek, Russian, and Brazilian Portuguese, are never presented with evidence that certain [GENDER] mismatches are ill-formed under ellipsis. Given this, the following question arises: are these learners exposed to other kind of data where a three-way asymmetry is detectable, and do these data then lead learners to categorize nouns into three classes that give rise to the ellipsis patterns as a side-effect? In other words, is there an independent trigger that is observable for the learner? Furthermore, if the availability of such a trigger is non-uniform for different populations of learners, can that explain the range of variation we observe regarding which class different noun pairs will fall into for different speakers? In this regard, the discussion in Sudo & Spathas 2020a of independent domains where three classes of nouns are detectable in Greek could provide a starting point for comparative work and the exploration of the kinds of relevant data that learners are exposed to during the acquisition process.

4.7.2 A contrast between NP ellipsis and predicate ellipsis?

As we discussed before in 4.6, Merchant 2014 reported a contrast between NP ellipsis and predicate ellipsis regarding the availability of [GENDER] mismatches in Greek—namely, Merchant reported that [GENDER] mismatches are banned altogether with NP ellipsis, but a tripartite asymmetry based on noun class membership arises under predicate ellipsis. However, later works like Sudo & Spathas 2016 and Alexiadou 2017 showed that the putative asymmetry reported by Merchant was not empirically supported, since the NP ellipsis examples that were presented had independent confounds rendering them all ill-formed. In a nutshell, the empirical picture seems to show that all ellipsis types in Greek showcase the three-way asymmetry (though see comments in footnote 132). In Brazilian Portuguese, the picture is similar, since Bobaljik & Zocca 2011 present a three-way asymmetry using predicate ellipsis, while Nunes & Zocca 2009 show that [GENDER] mismatches with NP ellipsis depend on noun class membership as well. Given this empirical landscape, then, it is important to ask ourselves at this juncture whether grammars *do* exist where ellipsis type makes a difference—do we observe variation in this regard? If so, how would we account for it?

It can be difficult to evaluate whether the literature on Spanish and [GENDER] mismatches provides unequivocal evidence that such a system is attested, given that most authors only present a subset of the crucial data. Taking partial reports from different sources and cobbling them together, of course, would not provide the evidence we need, given the existence of microvariation (see 4.7.1 above). Consider, for example, Picallo 2017, which makes the blanket statement that [GENDER] simply cannot mismatch under ellipsis, in contrast to [NUMBER]. The author only discusses the following NP ellipsis examples, noting that "*Gender alternation is not allowed under nominal ellipsis,* **in**

particular *if the nominal construction functions as an argument of predication*" (*emphasis*

mine; p. 6):[141]

(155)   *NP ellipsis examples adapted from Picallo 2017*
        a. *Juan admira   a su hermana pero Pedro envidia al   <hermano>
           Juan admires A his sister   but Pedro envies A.the.M brother
        suyo.
        his.M
        Intended: 'Juan admires his sister, but Pedro envies his (brother).'
        b. (*)? Juan admira a su hermano pero Pedro envidia a la <hermana> suya.
        Intended: 'Juan admires his brother, but Pero envies his (sister).'
        c. (*)? El   gato de pelo largo y   la   <gata> de pelo corto estaban en
               the.M cat.M of hair long   and the.F cat.F of hair short were   on
        la azotea.
        the roof
        Intended: 'The long-haired cat and the short-haired female cat were on the roof.'
        d. *Mientras María admiraba a las   actrices cómicas, Pedro alababa a
            while   María admired A the.F.P actresses comic   Pedro praised A
        los   <actores> trágicos.
        the.M.P actors   tragic
        Intended: 'While María admired the comic actresses, Pedro praised the tragic
                actors.'

Using this dissertation's terminology, the examples above are ill-formed mismatches with

Class I nouns (9)a-c and an ill-formed mismatch with a Class III pair, where the feminine

is in the antecedent (9)d.[142]

---

[141] On a similar note, Saab 2019: 538 provides only a *tío/tía* alternation (our Class I) and states: "*Although*

nouns differ as to how natural gender is morphologically represented, the ban on nominal ellipsis under

*gender mismatches remains constant **with some subtle differences among speakers***." (emphasis mine)

[142] A footnote in Picallo 2017 remarks that a reviewer commented that speakers favor the masculine in the

antecedent in examples like those provided. The author attributes this to the fact that masculine is the

unmarked grammatical gender in Spanish. It is impossible to know whether the reviewer meant that these

specific classes of nouns allow for an asymmetrical behavior (i.e., they are all Class III for some speakers),

or whether the reviewer made a more general claim about [GENDER] mismatches overall.

One cannot discern whether the grammar analyzed by Picallo represents a system where [GENDER] truly cannot mismatch at all (as is claimed in the main text) or a system where there is an asymmetry (of some sort) between NP ellipsis and predicate ellipsis. Furthermore, the data provided are consistent with the data presented in this chapter, since all these examples are judged ill-formed by my consultants and me. Put differently, since only a subset of the necessary data is given, it is impossible to evaluate whether the pattern illustrated by Picallo instantiates a grammar that could not be accounted for directly by the analysis presented here.

Similar concerns arise when evaluating other sources. For example, Donatelli 2019 mostly discusses predicate ellipsis examples. Only two sets of NP ellipsis examples are given, citing Kornfeld & Saab 2004 and Saab 2010a:[143] an ill-formed mismatch involving the Class I pair *tío/tía* 'uncle/aunt' and another involving the pair *padre/madre* 'father/mother' (see 4.5.1 above on Class IV and V nouns). Therefore, it is also impossible to assess whether the grammar analyzed by Donatelli 2019 is one where ellipsis type matters—i.e., one cannot tell if the three-way asymmetry reported in that work is exclusive to predicate ellipsis or not.

Saab 2008, however, presents a few examples which show that the grammar analyzed therein distinguishes between ellipsis types. The following are the crucial examples; first,

---

[143] Kornfeld & Saab 2004 only provide one example of [GENDER] mismatch under NP ellipsis, namely the *tío/tía* alternation. They use this example to state that [GENDER] cannot mismatch under NP ellipsis and attribute this observation to Depiante & Masullo 2001. Note, however, that Merchant 2014 cites Depiante & Masullo 2001's Spanish data to show the three-way asymmetry with predicate ellipsis. I have been unable to obtain the handout by Depiante & Masullo 2001.

Saab notes that a [GENDER] mismatch in NP ellipsis with inanimates is strongly
unacceptable. Furthermore, the same mismatch with a noun like *médico/médica* 'doctor'
(which we would place in Class II) is less deviant. Finally, a [GENDER] mismatch with
predicate ellipsis is perfectly formed (though see (152)c above from Saab 2004):

(156)   *Saab 2008: NP ellipsis versus predicate ellipsis (repeated from (125))*

    a. *El   manzano de Pedro es muy viejo, pero la   <manzana> suya
       the.M apple.tree of Pedro is very old   but the.F apple   its.F
       todavía es la   mejor.
       still   is the.F best
       Intended: 'Pedro's apple tree is very old, but its apples are still the best.'

    b. ?? El   médico de Pedro es viejo pero la   <médica> mía
       the.M doctor.M of Pedro is old   but the.F doctor.f my.F
       todavía es joven.
       still   is young
       Intended: 'Pedro's (male) doctor is old, but my (female) doctor is young.'

    c. Juan es maestro y   María también <es maestra>.
       Juan is teacher.M and María also   is teacher.F
       'Juan is a teacher and María is too.'

    d. María es maestra y Juan también <es maestro>.
       'María is a teacher and Juan is too.'

Saab 2008: 545 also provides ill-formed [GENDER] mismatches with pairs like *padre/madre*
'father/mother' (see our Class IV and V in 4.5.1) and with noun pairs denoting nobility
titles.

On the one hand, we can conclude from the data above that there exists *some*
asymmetry with different noun pairs under predicate ellipsis. On the other hand, if we take
(156)b above to be representative of a grammar where all NP ellipsis examples with a
[GENDER] mismatch are ill-formed, then the description in Saab 2008 does instantiate a
system where ellipsis type makes a difference. However, a full, controlled set of examples
involving all the relevant nouns is not found in Saab 2004 or 2008. As far as I can tell,
neither Saab 2004 nor Saab 2008 provide predicate ellipsis examples with kinship terms

like *tío/tía* (which would belong to Saab 2004, 2008's inflectional pairs and should behave akin to *maestro/a* in (156)b) or Class III nouns like *actor/actriz*.[144]

Moving on to another possible grammar exhibiting an asymmetry based on elliptical constructions, van Craenenbroeck & Merchant 2013 cite Masullo & Depiante 2004a and indicate that there exists an NP ellipsis versus predicate ellipsis asymmetry in Spanish. A claim is also made that this same asymmetry is also found in Brazilian Portuguese and Greek.[145] The Spanish examples given by van Craenenbroeck & Merchant 2013 are shown below—the first shows a well-formed [NUMBER] mismatch, while the second shows an ill-formed [GENDER] mismatch. Note that the noun pairs *tío/tía* and *abogado/abogada* are collapsed and the judgement given is the same for both noun pairs:[146]

(157)   *NP ellipsis in Spanish—van Craenenbroeck & Merchant 2013*
        a. Juan visitó a su { tío / abogado} y   Pedro visitó a los   { tíos / abogados}
           Juan visited A his uncle/lawyer.M and Pedro visited A the.P.M uncles/lawyers.M
           suyos.
           his.M.P
           'Juan visited his {uncle / lawyer} and Pedro visited his ({uncles / lawyers}).'

---

[144] Note however, that Saab 2010a: footnote 12 cites those two earlier works for the claim that *actor/actriz* alternations are symmetrically ill-formed under predicate ellipsis—I am unable to find such data therein.

[145] The assertion about Greek is based on a manuscript that predates Merchant 2014. The assertion about Brazilian Portuguese is puzzling, though, since the authors cite Nunes & Zocca 2009 in the paper. Recall that Nunes & Zocca show that [GENDER] *can* mismatch in NP ellipsis with Class II nouns (see (12)). As far as I can tell, then, the Spanish facts reported by van Craenenbroeck & Merchant 2013 are *not* identical to the report in Nunes & Zocca.

[146] van Craenenbroeck & Merchant 2013 cite a manuscript Masullo & Depiante 2004a. I have only been able to obtain Masullo & Depiante 2004b, which is a handout. Only the NP ellipsis data involving *tío/tía* 'uncle/aunt' and (*el) testigo / (la) testigo* 'witness' are in the handout; there is no example with *abogado / abogada* 'lawyer'.

b. * Juan visitó a su { tío / abogado} y   Pedro visitó a la { tía / abogada}
Juan visited A his   uncle/lawyer.M and Pedro visited A the.F aunt/lawyer.F
suya.
his.F
Intended: 'Juan visited his {uncle / lawyer} and Pedro visited his ({aunt / female
lawyer}).'

The authors also provide one example with the epicene noun *testigo* 'witness', which I

classified as Class II earlier. The example involves a feminine-masculine mismatch

(Masullo & Depiante 2004b do provide a masculine-feminine mismatch as well and judge

it unacceptable too):

(158)   *NP ellipsis in Spanish—van Craenenbroeck & Merchant 2013*
*El fiscal   interrogó a la   testigo del   crimen y   el juez
the prosecutor interrogate A the.F witness of.the crime   and the judge
interrogó   al   <testigo> del   robo.
interrogated A.the.M witness of.the robbery
Intended: 'The prosecutor interrogated the (female) witness of the crime and the
judge interrogated the (male) witness of the crime.'

While my consultants and I disagree with the judgements provided by these authors—the

only unacceptable example is the alternation of Class I *tío/tía* 'uncle/aunt' in (157)—what

is relevant for our purposes is that Masullo & Depiante 2004b do provide one example that

suggests that ellipsis type does matter for some speakers. They provide the following data

point with the epicene noun *(el) líder / la líder* 'leader'; contrast this example with (156):

(159)   *Ellipsis type seems to matter (adapted from Masullo & Depiante 2004b)*
María se   convirtió en una gran líder,   y   su hermano también <se.
Mary REFL converted in a   great leader.F and her brother.M too   REFL
convirtió en un gran líder>.
converted in a   great leader.M
'Mary became a great leader, and so did her brother.'

Masullo & Depiante suggest based on the above that *all* predicative nouns enter into the derivation lacking [GENDER] features.[147]

A puzzle arises, then, not just for the proposal in this dissertation, but for any account of [GENDER] mismatches under ellipsis that seeks to locate the explanation of the empirical picture in the identity condition. The analysis proposed in this chapter deals well with grammars like the Spanish analyzed here, Brazilian Portuguese (Bobaljik & Zocca 2011; Nunes & Zocca 2009), and Greek (Sudo & Spathas 2016, Alexiadou 2017), for which ellipsis type does not matter and the asymmetries that are attested can be explained via (i) a fine-grained decomposition of nominals, (ii) an identity condition predicated on featural non-distinctness and $\sqrt{\text{ROOT}}$ identity, and (iii) repair-by-ellipsis. All else being equal, ellipsis type should not matter, and the analysis proposed here captures a world where that is true.

While I leave as an open question what kind of account can capture the data we presented in 4.2 and 4.5 *in conjunction* with data where a distinction between NP ellipsis and predicate ellipsis makes a difference, one possibility comes to mind. Echoing the spirit of Merchant 2014's approach, it is possible that for speakers where ellipsis type matters, one of the constructions does not involve true ellipsis (i.e., it is not a surface anaphor in the Hankamer & Sag 1976 sense), but is instead a deep anaphor. In other words, suppose that NP ellipsis for Saab's consultants involves a surface anaphor that is subject to the identity

---

[147] This analysis is difficult to reconcile with the data in Depiante & Masullo 2001 (reported by Merchant 2014) showing the tripartite asymmetry in predicate ellipsis that is dependent on noun class membership. In other words, Masullo & Depiante (2004)b suggest an analysis that would predict that all mismatches in predicate ellipsis are well-formed, contra their previous reports.

condition proposed in this dissertation. Imagine, furthermore, that those speakers represent all nouns under analysis as Class I—i.e., the masculine is licensed by masculine $n_{i[-F]}$. We would thus derive the unavailability of [GENDER] mismatches via the identity condition proposed here.

In contrast, perhaps so-called predicate "ellipsis" for this set of speakers involves a null pronominal, which, of course, is not subject to the identity condition on ellipsis. Thus, a broader range of mismatches are available, and the availability is not governed by the featural decomposition of the nominals and the identity condition, but semantico/pragmatic considerations that would need to be established. Naturally, this proposal makes predictions, such that for those speakers, sub-extraction from gaps in predicate "ellipsis" should be impossible, at least in those cases where there is a mismatch in grammatical gender. I leave such an investigation for the future.

4.8 Back to [NUMBER]

In this section, I address a claim in the literature that [NUMBER] mismatches in certain configurations are, in fact, ill-formed under ellipsis (Picallo 2017).[148] If this claim were to

---

[148] Picallo 2017's claim appears to have an antecedent in Masullo & Depiante 2004b. Masullo and Depiante consider certain nouns like *bodas* 'wedding' and *noticias* 'news' to be plural even when denoting an atomic entity, and they claim that a number mismatch under ellipsis is ill-formed. They note, however, that these nouns can alternate with a singular version, especially in "less formal registers":

(i)   *Asistí   a la   boda   de Susana, pero no a las   de Pablo.
         attended to the.S wedding of Susan   but not to the.PL of Pablo
         Intended: I attended Susan's wedding, but not Pablo's (wedding).'

Their examples, however, exhibit the same problem that I will discuss below for Picallo 2017's data—they do not provide a controlled non-elliptical example. Those controls are just as unacceptable with the intended

be maintained, we would have to rethink the distribution of features in the nominal spine and consider how these examples inform the identity condition proposed in this dissertation. I will show, however, that the data provide no evidence for this claim, and I will therefore conclude that the asymmetry between [GENDER] and [NUMBER] regarding ellipsis is indeed absolute—whereas [GENDER] mismatches display an intricate behavior that varies depending on noun class membership, [NUMBER] mismatches are well-formed across the board.

Picallo 2017 may be seen to show that [NUMBER], in some cases, cannot mismatch under ellipsis. The data supporting this claim involve alternations with *pluralia tantum* nouns—these are syntactically plural nouns which nevertheless denote an atomic entity (for example, English 'scissors', '(reading) glasses'):

(160)   *Pluralia tantum nouns* (Picallo 2017)
        tijeras, gafas, pantalones, tenazas, ojeras
        'scissors, glasses, pants, pliers, bags (under one's eyes)'

As noted in Picallo 2017, certain speakers of Spanish accept the singular or plural version of these nouns without any meaning difference. For those speakers, the paper reports that a [NUMBER] mismatch is ill-formed under ellipsis.

My consultants and I allow for this alternation in nominals and agree with the judgments reported in the paper, given below as (161)a,c. The examples in (161)b,d are

---

reading, where the grammatically plural version of the noun in the target clause is meant to denote an atomic entity. Therefore, examples like (i) are irrelevant for the identity condition on ellipsis.

control examples showing that a [NUMBER] mismatch is allowed freely for nouns that are in the same semantic field but are not *pluralia tantum*:[149]

(161) *Pluralia tantum nouns under ellipsis*
    a. *Sus pantalones están limpios, pero el   <pantalón> tuyo   está arrugado
      her pants   are clean   but the.SG pants   yours.SG is   wrinkled
      y   sucio.
      and dirty
      Intended: 'Her pants are dirty, but yours are wrinkled and dirty.'

    b. Sus camisas están limpias, pero la   <camisa> tuya   está arrugada
      her shirts   are clean   but the.SG shirt   yours.SG is   wrinkled
      y   sucia.
      and dirty
      'Her shirts are clean, but yours is wrinkled and dirty.'

    c. *Saca   esta tijera   de   la mesa y   pon estas <tijeras> en el cajón.
      remove this scissors from the table and put these   scissors in the drawer
      Intended: 'Remove these scissors from the table and put these in the drawer.'

    d. Saca   esta engrapadora de   la mesa y   pon estas <engrapadoras>
      remove this stapler   from the table and put these   staplers
      en el cajón.
      in the drawer
      'Remove this stapler from the table and put these in the drawer.'

The ill-formed examples above are unacceptable only under *one* interpretation, where the plural nominal is intended to denote an atomic entity (see Picallo 2017, fn. 12). A context is given below showing this explicitly in a controlled manner:

(162) *Pluralia tantum nouns under ellipsis*
    Context: Marta has two pairs of pants. They are clean. You have one pair of pants. They are wrinkled and dirty.
    Sus pantalones están limpios, pero el <pantalón> tuyo está arrugado y sucio. (161)a

The example above is well-formed because the context makes clear that the nominal *pantalones* 'pants' in the antecedent does not denote an atomic entity. However, in a

---

[149] Picallo's reported judgments for the relevant examples are [(??)?] for (161)a and [(*)??] for (161)c. I provide * as a contrast to the controls in (161)b,d.

context where this nominal *is* meant to denote an atomic entity, the example is unacceptable:

(163) *Pluralia tantum nouns under ellipsis*
    Context: Marta has one pair of pants. They are clean. You have one pair of pants. They are wrinkled and dirty.
    *Sus pantalones están limpios, pero el <pantalón> tuyo está arrugado y sucio.
                                             (161)a

The example as reported by Picallo 2017 must be judged in a context like the above. The same sort of context is needed for the judgment in (160)c—the example is ill-formed only if the nominal in the ellipsis site is meant to denote an atomic entity.

Picallo 2017 proposes that the ellipsis examples just discussed show that there must be a [NUMBER] feature located between the categorizing n^0 (which Picallo assumes bears [GENDER]) and the √ROOT:

(164) *A low position for [NUMBER] (Picallo 2017)*
    [DET … [INFL … [NUM] … [n(Gender) … "Number" … √ROOT]]]

Under the identity condition on ellipsis assumed in Picallo 2017—which requires strict syntactic identity—the pattern with *pluralia tantum* nouns is derived because the ellipsis site in NP ellipsis is *n*P, which is non-identical to the antecedent (see Picallo 2017: 7).

However, baseline examples are not given in Picallo 2017 to demonstrate that the non-elliptical controls of the crucial data are well-formed. In other words, if the reported contrast arises due to a violation in the identity condition on ellipsis, then the non-elliptical controls should well-formed. This is not the case. Switching between a singular and plural version of the relevant nouns while intending both to denote atomic entities is ill-formed, even without ellipsis (165)b:

(165) *Pluralia tantum manipulation; non-elliptical controls (not in Picallo 2017)*

Context: Marta has 1 pair of pants. They are clean. You have one pair of pants. They are wrinkled and dirty.

a. *Sus pantalones están limpios, pero el pantalón tuyo está arrugado y sucio.

b. *Sus pantalones están limpios, pero el <pantalón> tuyo está arrugado y sucio.

see (161)a

Intended: 'Her pants are clean, but yours are wrinkled and dirty.'

c. Sus pantalones están limpios, pero los pantalones tuyos están arrugados
   her pants   are clean but the.PL pants   yours.PL are wrinkled
   y   sucios.
   and dirty

d. Sus pantalones están limpios, pero los <pantalones> tuyos están arrugados y sucios.
   'Her pants are clean, but yours are wrinkled and dirty.'

e. Su pantalón está limpio, pero el   pantalón tuyo   está arrugado y
   her pants is clean but the.SG pants yours.SG is wrinkled and
   sucio.
   dirty

f. Su pantalón está limpio, pero el <pantalón> tuyo está arrugado y sucio.
   'Her pants are clean, but yours are wrinkled and dirty.'

I add another example showing the same thing below; the non-elliptical version of the

crucial example is just as unacceptable:

(166) *Pluralia tantum manipulation with non-elliptical control*

Context: Sasha has 1 pair of scissors. They are on the table. Shea has one pair of scissors. They are on the chair.

a. *La   tijera de Sasha está sobre la mesa y las   tijeras de Shea
   the.SG scissors of Sasha is   on   the table and the.PL scissors of Shea
   también.
   too

b. *La tijera de Sasha está sobre la mesa y las <tijeras> de Shea también.
   Intended: 'Sasha's scissors are on the able and Shea's scissors are too.'

c. Las   tijeras de Sasha están sobre la mesa y las   tijeras de Shea
   the.PL scissors of Sasha are on   the table and the.PL scissors of Shea
   también.
   too

d. Las tijeras de Sasha están sobre la mesa y las <tijeras> de Shea también.
   'Sasha's scissors are on the table, and Shea's are too.'

333

e. La   tijera   de Sasha está sobre la   mesa y   la   tijera   de Shea
   the.SG scissors of Sasha are   on   the table and the.SG scissors of Shea
   también.
   too

f. La tijera de Sasha está sobre la mesa y la \<tijera\> de Shea también.
   'Sasha's scissors are on the table, and Shea's are too.'

The data above show that we can conclude that *something* bans the alternation of the morphologically singular and plural versions of these nouns mid-utterance, while intending for both versions of the noun to denote an atomic entity. However, whatever that is, it has nothing to do with the identity condition on ellipsis. As a result, these data cannot shed light on the proper formulation of the identity condition on ellipsis, nor do the ellipsis examples provide evidence that these nominals contain a [NUMBER] feature inside the ellipsis site in NP ellipsis that immediately dominates the √ROOT.[150]

To summarize, there is no evidence for the claim in Picallo 2017 that [NUMBER] cannot mismatch under ellipsis when *pluralia tantum* nouns are manipulated. Non-elliptical control examples show that the contrasts used to back the claim in Picallo 2017 have nothing to do with ellipsis.

---

[150] Omer Preminger suggests that speakers have two registers—one with the singular version of the nominals under discussion, and another with the *pluralia tantum* version. Ellipsis or no ellipsis, the issue here might arise from switching Spanish register mid-utterance. Note that the problem must be of a different nature from similar phenomena, such as the doublets described in Collins 2018, which are verbal pairs such as *dreamed/dream*t, *dived/dove*, *sneaked/snuck* that are accepted by individual speakers and could be used in the same utterance—e.g., 'I dreamed I would make it in life, and she dreamt the same thing.' (Omer Preminger p.c.).

4.9 Conclusion and open questions

At the outset of this chapter, I introduced an empirical puzzle which could inform the proper formulation of the identity condition on ellipsis, the main concern of this dissertation. In a wide range of languages, [GENDER] does not behave uniformly under ellipsis, in contrast to [NUMBER]. Noun pairs seem to fall into different classes whose mismatch behavior is unique: Class I disallows [GENDER] mismatches symmetrically; Class II allows them symmetrically; Class III exhibits an asymmetrical pattern, where only masculine-feminine mismatches are well-formed, but the inverse is not. I discussed how my final identity condition (repeated below) can provide insight into the derivations underlying this complex pattern—crucially, when coupled with independently justified approaches regarding (i) the decomposition of nominals and (ii) the possibility of repair-by-ellipsis of morphophonological gaps:

(167)   *Syntactic identity condition on ellipsis (final)*
   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.

Along the way, I proposed that derivations that comply with the identity condition above can *still be ill-formed*, given that the contents of an ellipsis site may involve a licensing violation between the nominalizing head *n* and a √ROOT. I followed Kramer 2015 in assuming that licensing violations are configurations that cannot be interpreted by the post-syntactic Encyclopedia. This aspect of my proposal has the potential to contribute to the debate on the status of different gaps (morphophonological or otherwise) under ellipsis.

I end this chapter by raising two questions that deserve consideration and that I intend to explore further in future work.

The first question is related to the interaction between [NUMBER] and [GENDER] under ellipsis. Consider that plural versions of masculine nouns in *all classes of nouns* permit a reading where a group is composed of individuals who use any pronoun set;[151] the plural versions of feminine nouns do not permit such a reading and only allow the interpretation where a group is composed only of individuals who use feminine pronouns (see (4), (5), and (6)).[152] However, the behavior of [GENDER] under ellipsis is the same regardless of any number manipulation—in other words, it is not the case that the [GENDER] mismatch asymmetries disappear when the manipulated nominals are plural: Class I mismatches are still symmetrically ill-formed, Class II mismatches are symmetrically well-formed, and so on. What this seems to tell us is that it is not the case that the addition of a [PLURAL] feature higher in the structure changes the featural composition of the *n*P, even though there is an interpretive difference between, say, *tío* 'uncle' and *tíos* 'uncles and aunts'. Grammatical gender and the semantic content related to gender, then, are dissociable even for nouns

---

[151] Of course, there exist efforts across the Spanish speaking world to avoid the use of these nominals in this way, given that such a usage is considered exclusionary. One possibility is to use a conjunction here—i.e., *abogados* y *abogadas*, though this could exclude those who use neither masculine nor feminine pronouns. Another alternative involves novel coinages like changing the last vowel to a "neutral" vowel like <e>: i.e., *abogades*.

[152] Bobaljik & Zocca 2011 claim that nobility terms differ in this respect, such that the plural of the masculine *princes* does not allow the mixed reading. They do mention in a footnote, however, that there were discrepancies in this judgment and a Spanish speaker allows for the plural *reyes* 'kings' to denote a mixed group. I agree with this judgement. It is also unclear to me whether Bobaljik & Zocca are claiming that other nouns that behave like nobility terms under ellipsis—i.e., *tíos/tías* 'uncles/aunts'—also do not allow a mixed reading. In Spanish, this is not so.

denoting humans. In other words, a mixed group behaves according to the grammatical properties of the class it belongs to—it seems that the expected behavior is not overridden by properties of the referents.[153]

A second issue concerns the *absolute status* of [GENDER] mismatches and the gradience of acceptability we observe regarding mismatches between the different classes of nouns. While the previous literature and my contribution in this chapter show that [GENDER] frequently displays an asymmetric pattern—where some mismatches are more acceptable that others based on noun class membership—[GENDER] mismatches are rarely perfect. Consider [NUMBER] mismatches as a point of comparison. Whereas [NUMBER] mismatches in ellipsis are *just as acceptable* as non-elliptical controls, authors consistently remark that the [GENDER] mismatches and asymmetries arising with them must be considered in relative terms (see my own comments in 4.2.2); i.e., mismatches with Class I nouns are terrible, whereas mismatches with Class II are acceptable (perhaps with a slight preference for masculine-feminine mismatches in comparison to feminine-masculine ones; see 4.7). When it comes to Class III, the masculine-feminine mismatches are acceptable in comparison to the feminine-masculine ones; the latter are simply impossible, while the former are better on a relative scale.

In light of this issue, consider my analysis of Class III mismatches, which crucially relied on repair-by-ellipsis. While examples involving gaps in verbal paradigms are perfect (see 4.3.3), the masculine-feminine mismatches with Class III are well-formed in comparison with the feminine-masculine mismatch. Why should this be the case, if in both examples—the one involving the verbal gap and the Class III masculine-feminine

---

[153] Thank you to Omer Preminger for discussion on this.

337

mismatch—(i) the identity condition is satisfied and (ii) repair-by-ellipsis is deployed? In carrying out syntactic work, researchers largely focus on the relative acceptability of examples, since the absolute status of sentences in isolation is rarely useful. However, it seems that any syntactic account of [GENDER] mismatches under ellipsis should also pay attention to the detectable, absolute differences in acceptability that are observed, in order to provide a fuller empirical picture and assess whether factors independent of the identity condition could be contributing to the patterns. The question that arises, then, is what these factors might be.

While we do not know at present the full range of factors that might be relevant, some recent work has attempted to address the kind of question posed above, allowing us to rule out some factors that conceivably could play a role. For example, Sprouse et al. 2020 report on an experimental investigation into mismatches in English ellipsis. The authors follow Bobaljik & Zocca 2011 in assuming that a mismatch is ill-formed symmetrically with pairs like *prince~princess*, while pairs like *actor~actress* display an asymmetric pattern, where only the masculine member of the pair is well-formed in the antecedent of a mismatch example. Sprouse et al. test whether this asymmetry in mismatch acceptability could be the result of a difference between these two types of pairs regarding the relative frequency in corpora between the masculine and feminine version of each pair, where the masculine is taken to be "unmarked". This frequency-based account (suggested in Haspelmath 2006) predicts that a mismatch would be worse the larger the frequency difference between the masculine member of a noun pair in comparison to the feminine. Sprouse et al., however, find no such frequency effect—in other words, a smaller or larger difference in frequency (based on corpus-counts) between the members of specific noun pairs does not correlate

with the size of the (un)acceptability effect observed in elliptical mismatches. These results suggest that frequency does not play a role in accounting for the clines of acceptability that the previous literature and the present chapter have remarked on for the languages we discussed. I leave for future exploration what other kinds of factors might be more promising to explain the acceptability asymmetries I have pointed out.

Chapter 5: Outlook

The proposal put forth in this dissertation has made empirical gains in different domains, yet there are areas that require further study and reevaluation, in particular if some empirical coverage appears to have been lost. In this chapter, I discuss possible extensions, as well as remaining challenges for the final identity condition proposed in this dissertation:

(1)   *Syntactic identity condition on ellipsis (final)*
      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.

I start in 5.1 by discussing apparent voice mismatches under sluicing in languages beyond Kaqchikel, focusing on two: Chamorro and Malagasy (Austronesian). I argue that neither language exhibits true voice mismatches under sluicing. Section 5.2 discusses Chung's generalization: the observation that ellipsis sites cannot contain "lexical items" that are not present in the antecedent. I zoom in on what kind of syntactic primitive counts as a "lexical item" for the purposes of such a condition, and propose that Chung's generalization follows from the second statement in the identity condition proposed here, which requires absolute √ROOT identity (1)b. "Vehicle change" phenomena are discussed in 5.3. I lay out why vehicle change is a problem for any syntactic identity condition, but outline a solution for a subset of the relevant data that would be compatible with the proposal in this dissertation. In section 5.4, I discuss cases where indexical switches lead to pronominal mismatches under ellipsis. I lay out a possible approach couched in Harley and Ritter 2002's featural decomposition of pronouns that could make these data compatible with the proposal here. Moving on to 5.5, I discuss Warner's Effects: the fact that auxiliaries in English appear to require absolute identity under ellipsis. Finally, in 5.6, I discuss an array of examples that

seem to suggest that a featural subset condition on ellipsis seems warranted, instead of the symmetric non-distinctness condition proposed in the present work. I show that either (i) the empirical picture is muddled by conflicting reports about the acceptability of certain configurations, or (ii) the ill-formed status of some data points may be orthogonal to the identity condition. Finally, 5.7 summarizes.

5.1 Voice mismatches beyond Kaqchikel? Chamorro and Malagasy

The availability of voice mismatches under sluicing has been reported in greatest detail for the Austronesian languages Chamorro (Chung 2006, 2013a) and Malagasy (Potsdam 2007).[154] A natural question to pose at this juncture is whether the syntactic identity condition proposed here is compatible with the Chamorro and Malagasy data.

In this section, I will argue that neither language shows evidence for a voice mismatch in the sense of our Kaqchikel discussion in chapter 2. In other words, "voice" mismatches in Chamorro or Malagasy are not well-formed merely because they satisfy featural non-distinctness. Rather, the data reported by Chung involving Antipassive-Active mismatches

---

[154] AnderBois 2008 also reports a voice mismatch between the antipassive and active voice in Yukatek Mayan. However, it is unclear whether the antipassive voice is a productive voice in the language, or whether antipassive morphology is a diachronic fossil (Scott AnderBois p.c.). If it is the latter, then there is no evidence for voice mismatches under sluicing in Yukatek. Drummond 2021 discusses apparent voice mismatches under sluicing in Nukuoro, a Polynesian-Outlier language from Micronesia but proposes that they instantiate a voice *match* and repair-by-ellipsis. Drummond proposes that the impossibility of certain argument structure alternations in Nukuoro sluicing argues against the proposal in this dissertation and in favor of strict identity, but it is unclear how the data are not ruled-out because they instantiate clashing *v* heads, in a manner similar to causative-inchoative alternations in English and other similar transitivity alternations (see chapter 3 here and Merchant 2013a).

in Chamorro can be analyzed (i) as repair by ellipsis (Chung 2006) or (ii) by positing that the "antipassive" is in fact an active, implicit argument construction (Chung 2013a). Turning to Malagasy, the mismatches reported in Potsdam 2007 seem to involve an agreement mismatch where one of the elements entering into the agreement relation is outside of the ellipsis site (see Pearson 2005). Given this analysis, the language does not exhibit voice mismatches of the Kaqchikel type either. The relevant examples are ruled in, however, by the requirement of the identity condition proposed here that only material properly contained within the ellipsis site must satisfy featural non-distinctness.

5.1.1 Chamorro

Chung (2006) highlights that a subset of voice mismatches under sluicing are allowable in Chamorro, an Austronesian language spoken in the Mariana islands. While Chamorro is an Austronesian language, Chung (1998, 2006, 2013a) argues that it does not possess a "Philippine"-style voice system (Erlewine et.al. 2017), in contrast to Malagasy and other languages in the family. Chamorro has three voices: active, passive, and antipassive. Chung 2013a:13-14 shows that Active-Passive mismatches are disallowed in Chamorro sluicing, which indicates that syntax plays a role in regulating the availability of ellipsis in the language. The picture becomes more complicated, however, when the antipassive voice is considered.

First, consider the examples below illustrating such a voice. The verbs here are intransitive, bear the morphological signal of the antipassive, and the theme is oblique:

(2)   *Chamorro AP* (Chung 2013a: 34)
    a. Kao mam-omoksai   si   Juan månnuk pa ngånga'?
        Q   AGR.AP-raise.PROG UNM Juan chicken or duck
        'Is Juan raising chickens or ducks?'

b. Mungnga mam-atcha   ni iyu-n   ottru tåotao.
   don't   AGR.AP-touch OBL possession-L other person
   'Don't touch other people's things.'

c. Mang-guaiya hao as   Pedro.
   AGR.AP-love   you OBL Pedro
   'You love Pedro.'

In order to build the relevant mismatches under sluicing, it is crucial to establish first that the oblique argument of the antipassive cannot undergo *wh*-extraction. This is shown below:

(3)   *Oblique in Chamorro AP cannot undergo wh-movement* (Chung 2013a: 35)

   a. *Håfa   na klåsi-n månnuk mam-omoksai   gui'?
      what? L sort-L   chicken AGR.AP-raise.PROG he
      Intended: 'What sort of chickens is he raising?'

   b. *Håyi mang-guaiya hao?
      who? AGR.AP-love you
      Intended: 'Who do you love?'

A *wh*-remnant can therefore be manipulated to construct a sluicing mismatch where the antecedent is in the antipassive voice and the target clause is (presumably) in the active voice. This type of example is well-formed:

(4)   *AP-Act mismatch is possible in Chamorro*

   a. Mam-omoksai månnuk, lao ti   ta   tungu' håfa   na klåsi.
      AGR.AP-raise   chicken but not AGR know what? L sort
      'He is raising chickens, but we don't know what kind.'   (Chung 2006: 78)

   b. Mang-guaiya si   Julia, lao ti   hu   tungu' håyi.
      AGR.AP-love UNM Julia but not AGR know who?
      'Julia loves (someone), but I don't know who.'   (Chung 2013a: 35)

The examples in (4) need to be accounted for, since their well-formedness might imply at first glance that they comply with whatever identity condition regulates ellipsis. Three analytical possibilities present themselves for the content of the ellipsis site here; only the first of these would be a problem for the identity condition proposed in this dissertation.

The first analytical possibility is that these examples instantiate a genuine Voice^0 mismatch, where the *wh*-remnant ensures that the target clause is in the active voice. Under this interpretation, $\text{Voice}_{\text{ACTIVE}}$ in the ellipsis site would clash with Voice_{AP} in the antecedent, but the example would be well-formed. In a nutshell, Chamorro would be the empirical opposite of Kaqchikel, where examples of this kind are banned, and sluicing in the language would be a real empirical challenge to any syntactic identity condition. Whereas this interpretation of the Chamorro data in (4) must be considered for the purposes of this subsection, neither Chung 2006 nor Chung 2013a propose such an interpretation of the facts.

Rather, Chung (2006) advances a second interpretation—namely, the mismatch in examples like (4) is an illusion because the *wh*-remnant has not, in fact, ensured that the ellipsis site is a clashing Voice^0. Instead, these examples instantiate an AP-AP *match*, where the oblique argument of an antipassive clause has been *wh*-extracted. Even though this extraction is illicit in non-elliptical contexts, sluicing repairs the problem (presumably, an island violation).[155]

A third approach is advanced in Chung 2013a, which rejects the island repair analysis because the resulting derivation would violate the Case component of the identity condition proposed in that work.[156] Instead, Chung (2013a) argues that examples like (4) create the

---

[155] In contrast to our discussion of the Kaqchikel AP-Active mismatch (which is ill-formed; see chapter 2), one would be forced to posit that in Chamorro, extracting the theme of the antipassive creates a PF problem, not a syntactic problem.

[156] Chung 2013a's identity condition is reproduced below:

illusion of a mismatch, but for a different reason than that proposed in Chung 2006: the AP construction in Chamorro is not an antipassive in the strictest sense, but is rather an implicit argument construction whose syntax is built up from the same components as the active voice.[157] Chung argues that the examples in (4) involving an antipassive "voice" in the antecedent should be analyzed in the same manner as examples like the English below:

(5)   She's **reading**, but we can't figure out what <she's reading>. (Chung 2013a: 39)

(i)   *Limited syntactic identity in sluicing* (Chung 2013a: 30)

   a. *Argument structure condition*: If the interrogative phrase is the argument of a predicate in the ellipsis site, that predicate must have an argument structure that is identical to that of the corresponding predicate in the antecedent clause.

   b. *Case condition*: If the interrogative phrase is a DP, it must be Case-licensed in the ellipsis site by a head identical to the corresponding head in the antecedent clause.

Assessing the strengths and weaknesses of the proposal above would take us too far afield. However, a couple of notes are in order. First, the identity condition here is specific to sluicing. Second, this identity condition has no purchase on the different status of TP-level mismatches in English and Spanish, which I discussed in chapter 3. Third, the argument structure condition can be restated as a requirement within the verbal domain that heads match one-to-one (see the discussion of Rudin 2019's condition in chapter 2 and 5.2 below) and is thus challenged by the Kaqchikel facts in chapter 2. Finally, convincing arguments against employing the Case condition in (i)b for the relevant examples discussed in Chung 2013a have been put forward in the literature (see Barros 2014, Thoms 2015, Rudin 2019). If the empirical support used in Chung 2013a for the specific Case condition above are not strong, then such a condition should not be used to justify rejecting Chung's earlier 2006 analysis of repair by ellipsis for data like (4).

[157] Chung (1998: 38-39) notes that only a subset of transitive verbs in Chamorro participate in the active-antipassive alternation. This may suggest that there are independent reasons to argue that the "antipassive" in the language is not a wholly productive voice.

There is no clash of Voice heads under this analysis, then—the featural make-up of the "antipassive" voice in Chamorro is not different from the active voice. Therefore, examples of the kind seen in (4) are not problematic for the identity condition defended here.

To summarize, Chamorro examples appear at first glance to be a challenge for the identity condition defended in this dissertation. Specifically, the examples in (4) seem to instantiate well-formed examples where a Voice_{AP} in an antecedent clashes with a Voice_{ACT} in the ellipsis site, a configuration that I showed is ill-formed in Kaqchikel. However, it seems that the data do not instantiate a true voice mismatch. Instead, the examples can be analyzed as an instance of our now-familiar repair by ellipsis of an island violation (Chung 2006). Alternatively, the "antipassive" in Chamorro may not be a true antipassive. Rather, it may be an implicit argument construction whose syntactic building blocks are the same as an active transitive (Chung 2013a).

5.1.2 Malagasy

Potsdam 2007 shows that Malagasy allows for voice mismatches under sluicing. The question that interests us here is whether the Malagasy data are comparable to the Kaqchikel data we analyzed in chapter 2. However, I will show in this subsection that the Malagasy data are not akin to the Kaqchikel data: the "voice" mismatches in Malagasy can be plausibly analyzed as instances of agreement mismatches that have nothing to do with voices like active, passive, and antipassive. Therefore, the Malagasy examples are not cases of extraordinarily well-formed clashes between Voice^0 heads, and the data are ruled in by the identity condition defended here.

Malagasy is an Austronesian language with a so-called "Philippine"-style voice system (Erlewine et.al. 2017). The language distinguishes between three "voice" inflections on

verbs, which alternate depending on which argument functions as a *trigger*. The trigger occupies a peripheral position in the clause. If the external argument is the trigger, the Actor Trigger (AT) voice is used; if the internal argument is the trigger, then the Theme Trigger (TT) voice is used. Finally, if certain adjuncts function as the trigger, then the Circumstantial Trigger (CT) voice appears:

(6)   *Malagasy "voices"* (Pearson 2012)
    a. **Namono** ny akoho tamin' ny antsy **ny mpamboly**.
      PST.AT.kill DET chicken with   DET knife **DET farmer**
      'The farmer killed the chicken with the knife.'

    b. **Novonoin'** ny mpamboly tamin' ny antsy **ny akoho**.
      DST.TT.kill DET farmer   with   DET knife **DET chicken**
      'The farmer killed the chicken with the knife.'

    c. **Namonoan'** ny mpamboly ny akoho   **ny antsy**.
      PST.CT.kill DET farmer   DET chicken **DET knife**
      'The farmer killed the chicken with the knife'

What is most relevant for our purposes is that *wh*-phrases must be triggers. In the transitive examples below, an external argument *wh*-phrase requires the AT voice, while an internal argument *wh*-phrase requires the TT voice:

(7)   *Malagasy wh-question formation* (adapted from Potsdam 2007: 581-583)
    a. **Iza** no **mividy** ny osy?
      who PRT buy.**AT** the goat
      'Who is buying the goat?'

    b. ***Iza** no **vidina** ny osy?
       who PRT buy.**TT** the goat
      Intended: 'Who is buying the goat?'

    c. **Inona** no **vidin'** i Soa?
      what PRT buy.**TT** Soa
      'What is being bought by Soa?'

    d. ***Inona** no **mividy** i Soa?
       what PRT buy.**AT** Soa
      Intended: 'What is being bought by Soa?'

Potsdam 2007 shows that the Malagasy voice alternations shown above can mismatch in sluicing. Consider the examples below, which show an AT-TT and TT-AT mismatch. Note

that the way to force the mismatch is via the manipulation of the *wh*-remnant; using an external argument remnant forces AT voice in the ellipsis site, while an internal argument remnant forces the TT voice:

(8)   *Malagasy AT-TT voice mismatch* (adapted from Potsdam 2007: 584-85)
   Nandoko zavatra i Bao fa   hadinoko   hoe **inona** <… **nolokoin'** …>.
   paint.**AT** thing Bao but forget.TT.1SG COMP **what**   paint.**TT**
   'Bao painted something but I forget what.'

(9)   *Malagasy TT-AT voice mismatch* (adapted from Potsdam 2007: 584-85)
   Nangalarin' ny olona ny fiarako fa tsy fantatry ny polisy hoe
   steal.**TT**   the person the car.1SG but NEG know   the police COMP
   iza <… **nangalatra** …>
   who   steal.**AT**
   'My car was stolen by someone but the police don't know who.'

The question at hand is the following: are these mismatches of the same nature as the voice mismatches we have discussed in English and Kaqchikel (e.g., Active-Passive, AF-Active, etc.)?

Some authors have indeed assumed that the AT and TT voices instantiate active and passive respectively. In other words, the trigger is taken to be the grammatical subject (see e.g., Keenan 1976 and Pearson 2005 for discussion). If this view were correct, Malagasy would be a language that extraordinarily allows for voice clashes that are unattested in sluicing in other languages. From a theoretical perspective, then, Malagasy would be a challenge to any approach that assumes that some degree of syntactic identity regulates the availability of ellipsis. This is the conclusion drawn in Potsdam 2003, which argues that the Malagasy voice mismatches present one piece of evidence that the identity requirement

on ellipsis is purely semantic and based on mutual entailment (Merchant 2001's e-GIVENness), as opposed to being syntactic.[158]

However, the mapping of AT and TT to active and passive voices respectively has been argued to be incorrect. Pearson 2005 argues that the Malagasy "voice" morpheme instantiates the spell-out of an indirect agreement relation with the trigger (see also Rackowski and Richards 2005). More specifically, the trigger occupies an A'-position, the specifier of Top(ic)P, and binds an operator within the predicate phrase. The verb agrees with the operator and the spell-out of this agreement is the relevant voice morpheme—AT voice if the trigger binds an operator merged as an external argument; TT if the trigger binds an operator merged as an internal argument. This is schematized below:

(10)   *Malagasy "voices" are the spell-out of agreement* (Adapted from Pearson 2012)
         [PredP   ***Op*$_i$ V** … t$_1$ …] Trigger$_i$
                  Agreement

In a nutshell, then, the Malagasy voice mismatches under sluicing involve an instance of agreement mismatches, as opposed to a voice mismatch of the sort discussed in chapter 2. In the example below, the trigger is the external argument, leading to AT agreement, whereas the trigger in the target clause is the internal argument—leading to TT agreement in the ellipsis site:

(11)   *Malagasy AT-TT agreement mismatch* (adapted from Potsdam 2007: 585)
         Nandoko zavatra i Bao fa hadinoko   hoe **inona**$_i$ < **Op**$_i$ no **nolokoin'**
         paint.AT thing   Bao but forget.TT.1SG COMP what   PRT paint.TT
         i Bao t$_i$ >.
         Bao
         'Bao painted something but I forget what.'

---

[158] Potsdam (2003) cites Merchant 2001 for the observation that voice *cannot* mismatch in English sluicing—a problem for a semantic identity condition; see chapter 2—but sets the issue aside.

The interpretation of the Malagasy facts just outlined is assumed by Potsdam (2007), who remarks that the apparent voice mismatches are inconclusive and cannot adjudicate between different formulations of the identity condition (see also Chung 2006).

An issue that arises as we assess the agreement analysis of "voice" in Malagasy is that the structure of *wh*-questions assumed by Pearson (2003)—who defends this analysis—and Potsdam (2007)—who accepts that the analysis is plausible—is not the same. Potsdam argues at length that *wh*-questions in Malagasy are pseudoclefts, as opposed to arising via regular *wh*-movement or a cleft-structure; consider the contrasting analyses below (for arguments against the cleft-analysis, see Potsdam 2007: 599-609):[159]

(12)   *Malagasy wh-question (Potsdam 2007: 594)*
         iza no mividy ny osy?
         who PRT buy   the goat
         'Who is buying the goat?'

---

[159] A significant question arises at this juncture—if pseudo-clefts truly underly Malagasy sluices, is the identity condition proposed in this dissertation satisfied by such sluices? Potsdam 2007 argues, in fact, that the Malagasy data are strong evidence against a strict requirement based on syntactic isomorphism. I cannot go into this question in detail, since it would be impossible to do justice to the substantial body of literature that argues that a cleft(-like) structure must underlie sluicing in at least a subset of configurations, even in languages like English (see van Craenenbroeck 2010a, Barros 2014, Vicente 2019 for discussion). However, the following point is important: my proposal in this dissertation would seem to correctly rule-in a cleft(-like) structure underlying the sluicing examples where such a structure must be posited, if we assume that the difference between a cleft(-like) and a non-cleft antecedent lies in the cleft(-like) ellipsis site containing *additional structure*. In a nutshell, the additional structure in the ellipsis site would not clash with any structure in the non-cleft antecedent, satisfying featural non-distinctness. If this is on the right track, then the argument from cleft(-like) sluicing against a syntactic identity condition vanishes (i.e., Potsdam 2007's proposal that Malagasy sluicing argues against syntactic isomorphism is correct, but Malagasy sluicing does *not* argue against the non-distinctness approach advanced here).

(13)   *Malagasy wh-question derivations (Potsdam 2007: 594)*
  a. *wh-movement derivation*
     [$_{CP}$ iza$_i$ [$_{C'}$ no [$_{FP}$ [$_{PredP}$ mividy ny osy] t$_i$]]]
          who   PRT   buy   the goat
     Literally: 'Who buys the goat?'
  b. *pseudocleft derivation*
     [$_{FP}$ [$_{Pred}$P iza] [$_{DP}$ Op$_i$ no   mividy ny osy t$_i$]]
               who   PRT buy   the goat
     Literally: '(One) that buys the goat is who?'

Let us assume, then, that the interpretation of "voice" in Malagasy as *agreement* is tenable.

We then need to explain *how* the identity condition on ellipsis is satisfied even though there

exists an agreement mismatch in Malagasy—it does not suffice to merely state that

agreement mismatches are tolerated under ellipsis, since that is the fact in need of an

explanation. Consider again in this regard the condition that I defended in this dissertation:

(14)   *Syntactic identity condition on ellipsis (final)*
  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.

In a nutshell, the agreement ("voice") mismatches in Malagasy sluicing can be accounted

for if the pseudo-cleft derivation in the target clause instantiates a predication relation that

involves Case-agreement between the *wh*-remnant and elements within the ellipsis site, and

agreement is the result of feature sharing (Pesetsky & Torrego 2001, 2007, Frampton &

Gutmann 2000, 2006, Chung 2013b). The Malagasy data can thus be analyzed along the

same lines as person mismatches in Spanish (chapter 2), since the features that are shared

among the elements entering into the agreement relation are not properly contained within

the ellipsis site.

To summarize, I conclude that Kaqchikel instantiates the first case in the literature where a bona-fide mismatch in Voice is attested under sluicing, mismatches that I argued comply with the identity condition proposed here.

5.2 Chung's generalization follows from √ROOT identity

One of the strongest arguments for a syntactic component to the identity condition on ellipsis comes from ill-formed data that took center stage in Chung 2006 (though relevant data can already be found in Rosen 1976). I will argue that those data can be subsumed under the requirement of √ROOT identity (1)b.

Consider first the examples below, adapted from Chung 2006:

(15)   a. She read something, but we're not sure by which author.
        b. She read something, but we're not sure [by which author]$_1$ <she read something t$_1$>.
        c. *She read, but we're not sure by which author.
        d. She read, but we're not sure [by which author]$_1$ <she read **something** t$_1$>

What seems to be causing the ill-formed status of (15)c is that there is an element in the ellipsis site, namely *something*, that has no correlate in the antecedent (15)d.

Moreover, consider the data in (16). Recall first that English is a P-stranding language—i.e., *wh*-movement need not pied-pipe a preposition. As expected, given that English is a P-stranding language, a preposition can be stranded under the merger type of sluicing (Merchant 2001, Chung et al. 1995, 2011; recall our discussion in chapter 3). However, Rosen 1976 showed already that a preposition cannot be stranded under the *sprouting* type of sluicing ((16)a,c from Rosen 1976):

(16)   *No P-stranding under sprouting*
        a. I'm getting involved, and I don't know exactly what in.
        b. I'm getting involved, and I don't know exactly [what in]$_1$ <I'm getting involved t$_1$ >.
        c. *I'm getting involved and I don't know exactly what.

d. I'm getting involved and I don't know exactly what$_1$ <I'm getting involved [**in** t$_1$] >.

e. Crystal was dancing, but I don't know who with.

f. Crystal was dancing, but I don't know [who with]$_1$ <Crystal was dancing t$_1$ >.

g. *Crystal was dancing, but I don't know who.

h. Crystal was dancing, but I don't know who$_1$ <Crystal was dancing [**with** t$_1$]>.

In a similar manner to the ill-formed (15)c, the issue seems to be that there is an extra element in the ellipsis sites in (16)c,g —in these instances, a preposition *in* or *with*.

Data like these led Chung to propose that there is a lexico-syntactic requirement on ellipsis. This is sometimes called the "No New Words Condition" (see Merchant 2013b, who provides a more formalized version). Even though (17) below has the flavor of being an analysis, it has often been referred to as Chung's generalization (see e.g., Barros 2014):[160]

(17)   *Chung's generalization (Chung 2006)*
      Every **lexical item** in the numeration of the sluice that ends up (only) in the elided IP must be identical to an item in the numeration of the antecedent CP.

I bold the term "lexical item" above, since establishing what counts as such will be relevant in the discussion hereafter.

One implementation of Chung's generalization within a single, unified identity condition is Rudin 2019's proposal. Recall that Rudin's proposal requires that every element inside the ellipsis site (that is within the eventive core) must find a structure-matching correlate in the antecedent (see chapter 3). In cases like (15)c and (16)c,g, the identity condition is violated, since there are elements within the ellipsis site that have no

---

[160] Occasionally, only the fact that P-stranding isn't allowed under sprouting (even in P-stranding languages) is referred to as Chung's generalization (see Anand et al. 2021). Chung's observation, however, goes beyond stranded prepositions with no correlate, as shown by the data in (15).

such correlate in the antecedent. However, we saw in chapter 3 that the key components of Rudin's proposal—that only elements within the eventive core enter into the calculation of the identity condition—cannot be correct. Thus, Chung's requirement cannot be wholly subsumed under the identity condition put forth in Rudin 2019. Since the cleaving apart of the clause into two domains is not supported—i.e., the eventive core is not special in this regard and the satisfaction of identity must be calculated over the entire ellipsis site—then Rudin's strict head-by-head matching requirement would deliver the wrong results (recall for instance the existence of polarity, voice, and grammatical gender mismatches; see below). The question that arises at this juncture is whether Chung's generalization follows from the identity condition proposed in this dissertation. I believe that the crucial issue is determining what counts as a *lexical item* in evaluating (17)—in other words, what the nature is of the elements within ellipsis sites that must match identically with an element in the antecedent.

To answer this question, let us first go over what could *not* fall under the umbrella of Chung's lexical items. Recall the examples in chapter 3 showing that polarity mismatches are well-formed in sluicing:

(18) *Polarity mismatch*
      I don't think that California **will comply**, but I don't know why <California **won't comply**>.

I argued that examples like the above are ruled in by the identity condition proposed in this dissertation because there is no clash between featurally specified Σ^0 heads—specifically, I analyzed (18) as a configuration where the ellipsis site contains $\Sigma_{[+NEG]}$ and the antecedent (a simple declarative) has no ΣP. If that proposal is on the right track, then it cannot be the case that a featurally specified head like $\Sigma_{[+NEG]}$ is the kind of element that is relevant for

Chung's generalization—the proposal here necessitates that there be a $\Sigma_{[+\text{NEG}]}$ in the ellipsis site that has no identical correlate in the antecedent.

The analysis of the well-formed AF-Active voice mismatches in Kaqchikel is instructive in this regard as well (see chapter 2). My interpretation of those examples leads us to conclude that a featurally specified Voice^0 head cannot be the kind of "lexical items" that Chung's generalization refers to either:

(19) *AF-Active mismatch*
      Xa   xe   ri   ma Pedro **x-Ø-loq'-o**   ri   kotz'i'j. Aw-etaman   ankuchi
      EMPH only DET CLF Pedro **COM-B3S-buy-AF** DET flower A2S-know-PERF where
      <x-Ø-u-löq'   wi>?
       COM-B3S-A3S-buy FP
      'Only Pedro bought the flowers. Do you know where?'

We concluded that examples like the above are ruled in by the identity condition proposed in the present work because they satisfy featural non-distinctness. While there is a Voice$_{\text{ACT}}$ in the ellipsis site, there is no VoiceP at all in the antecedent—recall that in an AF clause, VoiceP was removed by Exfoliation. Given that AF-Active mismatches are well-formed—even though Voice$_{\text{ACT}}$ in the ellipsis site does not have an identical correlate in the antecedent—we must conclude once more that featurally specified heads (in this case Voice^0) seem to be outside the scope of Chung's generalization in (17).

Finally, let us recapitulate the proposed analysis for a subset of the [GENDER] mismatches discussed in chapter 4. If my proposal is on the right track, there exist well-formed mismatches in that empirical domain in which a featurally specified head in the ellipsis site has no identical correlate in the antecedent. In these examples (Class II masculine-feminine mismatches), featural non-distinctness is satisfied because there is a mismatch between a default n in the antecedent (a nominalizing head that does not bear a [GENDER] feature) and a head $n_{i[+\text{F}]}$ in the ellipsis site:

(20)　*[GENDER] mismatch well-formed; Brazilian Portuguese (Nunes & Zocca 2009)*
　　a. O　João visitou o　　**médico** dele　e　　o　Pedro visitou **a**　　dele.
　　　 the João visited the.M **doctor.M** of.his and the Pedro visited **the.F** of.his
　　　 'João visited his doctor and Pedro visited his (female) doctor.'
　　b. … médico … a < [$_{nP}$ [$\textbf{\textit{n}}_{\text{i[+F]}}$ √MÉDIC]] > …

One last time, what this means is that an ellipsis site *can* contain a featurally specified head (in this case, *n* bearing a [GENDER] feature) that has no identical correlate in the antecedent. We observe, then, that featurally specified heads like $\Sigma_{\text{[+NEG]}}$, Voice$_{\text{ACT}}$, and $n_{\text{i[+F]}}$ do not fall under the "lexical items" in Chung's generalization. Let us generalize from this observation and conclude that all featurally specified heads are excluded from Chung's generalization. Put differently, the identity condition on ellipsis does not require that featurally specified heads in the ellipsis site have an identical correlate in the antecedent— instead, what is relevant is that there be no clashing head in the antecedent (a violation of non-distinctness), as I have argued throughout this dissertation.

What, then, is the kind of syntactic primitive that *does* fall under (17)? The natural candidate, and perhaps the only logically *possible* candidate (given the theoretical assumptions adopted in this dissertation), is √ROOTs. I propose, then, that Chung's generalization is derived by the √ROOT identity statement of the identity condition proposed in this dissertation (21)b, repeated below:

(21)　　*Syntactic identity condition on ellipsis (final)*
　　　 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.

Attributing Chung's generalization to (21)b straightforwardly derives the ill-formed status of an example like (15)c—the extra nominal 'something' in the ellipsis site contains a √ROOT for which there is no match in the antecedent.

What about prepositions, though, which we observed cannot be stranded under sprouting (16)? I follow the proposal in Deacon 2014, 2017 that adpositions contain a √ROOT that is categorized by a p^0 head. Consider as some evidence for this approach the categorical alternations of a √ROOT that can surface in nominal, verbal, adjectival, or adpositional frames; data in (22)a-c from Deacon 2017:[161]

(22)   *√ROOT categorized by n, v, a, p*
  a. The donors have an **in** with the senator.
  b. The villagers **inned** the marshes.
  c. The subtle Fiends, though **inly** stung with anger and disdain, dissembled.
  d. I am involved **in** a scandal.

Deacon proposes that there exists a categorial p^0 head that allows for a √ROOT to be used in an adpositional frame. If Deacon's proposal is on the right track, then we have some tentative independent support for the proposal that what goes wrong in examples like (16)c,g is that there is a √ROOT in the ellipsis site for which there is no match in the antecedent.

One might worry that this proposal is difficult to reconcile with the fact that no kind of preposition, even the most semantically bleached, can be stranded under sprouting:

(23)   a. Katya is jealous of someone, but I don't know who.
  b. Katya is jealous, but I don't know of who.
  c. *Katya is jealous, but I don't know who.
  d. Katya is jealous, but I don't know who$_1$ <Katya is jealous [**of** t$_1$]>.

The worry here might be that we would need to assume that even a preposition like *of* in the example above contains a √ROOT component. Though the issue is worthy of more discussion, I do not consider this to be an obstacle. Two comments are in order. First, *of*

---

[161] For a discussion on whether there exists a categorizing a^0 head, see Mitrović and Panagiotidis 2020.

certainly has non-vacuous semantics in other contexts, such as in its partitive use. Second, there are many elements like the nominals 'stuff', 'thing', or the verbs 'happen', etc., that are close to being semantically vacuous, but which by hypothesis are composed of a √ROOT and a categorizing head. I maintain, then, that deriving Chung's generalization from the √ROOT identity requirement is a promising area for further exploration.

To summarize this section, I discussed the data that led to Chung 2006's generalization. In a nutshell, there cannot be extra elements (of a particular nature) inside of ellipsis sites. Given the empirical results from previous chapters, I discussed how Chung's generalization cannot pertain to featurally specified heads. I proposed, thus, that Chung's generalization can be subsumed under the requirement of absolute √ROOT identity. I sketched out a proposal in the literature (Deacon 2017a,b) that adpositions contain a √ROOT component, which would entail (correctly) that they cannot be stranded under sprouting, given that such a configuration violates strict √ROOT identity.

5.3 Vehicle change

Perhaps the most challenging empirical domain for any identity condition that assumes a syntactic component is the phenomenon known as "vehicle change" (Fiengo & May 1994). I will focus here on pronoun/name vehicle change. Consider well-formed ellipses like the following:

(24)   *Vehicle change*
       They gave Trixie$_1$ the crown, though she$_1$ thought they wouldn't.

What is inside the ellipsis site here? The R-expression *Trixie* cannot be inside the ellipsis site, since that would result in a Principle C violation, and the example would be ruled-out, contrary to fact (25)a. Instead, it seem that there is a pronoun in the ellipsis site (25)b:

(25)   *Vehicle change*
    a. *They gave Trixie₁ the crown, though she₁ thought they wouldn't <give **Trixie₁ the crown>** .   *Principle C violation*
    b. They gave Trixie₁ the crown, though she₁ thought they wouldn't <give **her₁** the crown> .

Vehicle change has come to be the label for examples like the one above, where it seems that it is impossible to posit strict syntactic isomorphism between the ellipsis site and the antecedent.

The question that concerns us here is the following: if indeed the content of the ellipsis site is as specified in (25)b, how does such a configuration comply with the identity condition proposed here, or, for that matter, any identity condition predicated on syntax? In this regard, note that vehicle change phenomena constitute one of the key empirical domains that support semantic identity conditions like Merchant 2001's e-GIVENness.

To illustrate why a principled solution to this problem is difficult for any syntactic approach to the identity condition, it is useful to consider how the proposal in Rudin 2019—which I examined in chapter 3—handles cases like these. Recall Rudin's condition; the key aspect to focus on here is the definition of correlate in (27)b:

(26)   *Syntactic Condition on Sluicing (final)* (Rudin 2019: 269)
    Given a prospective ellipsis site E and its antecedent A, non-pronunciation of the phonological content associated with any head h ∈ E is licit if at least one of the following conditions hold
        a. h did not originate within E's eventive core
        b. h has a **structure-matching correlate** i ∈ A.

(27)   *Correlate* (Rudin 2019: 264)
    A node n can be a correlate for a head h iff at least one of the following conditions holds:
        a. n is a head and n and h are tokens of the same lexical item
        b. **n is coindexed with h**

In a nutshell, an R-expression counts as a correlate for a co-indexed pronoun by stipulation (see Rudin's footnote 23, where this seems to be acknowledged)—in other words, the proposal stipulates that an example like (25)b satisfies the identity condition.

I bring up this discussion in order to highlight the difficulty of accounting for vehicle change phenomena in a principled way—it is not my intention to claim that my identity condition *as is* can handle the facts better than Rudin's approach.[162] The question we are faced with at this juncture, though, is whether there is a way to reconcile vehicle change phenomena with the (independently arrived at) conclusion that syntax is a part of the identity condition, but featural *non-distinctness* is required.

In light of this, the discussion in Murphy 2016 might point towards a solution. In arguing for a subset condition on ellipsis (more on this below in 5.6), Murphy brings up the proposal in Saab 2015 (which follows Nunberg 1999 and Elbourne 2005) that pronouns are D elements with an index that, in the narrow syntax, stand in a proper subset relation with names, which are DPs bearing an index (note that Saab's proposal is a direct descendant of Oku 1998, who similarly proposed that pronouns are a featural subset of names). Under such an approach, it would be possible to pursue an analysis wherein an example like (24) is well-formed because having a pronoun in the ellipsis site (i) obviates a Principle C violation and (ii) satisfies the featural non-distinctness statement of the identity condition proposed here:

(28)   *Vehicle change*
       They gave **Trixie**$_1$ the crown, though she$_1$ thought they wouldn't <give **D**$_1$ the crown> .

---

[162] In other words, it is fair to note that a stipulation might need to be added to the condition proposed here in order to cover examples involving vehicle change.

Even though many questions remain—e.g., how pronouns acquire their φ-features under this approach—I leave this suggestion here as an issue for future exploration.[163]

5.4 Indexical/pronoun mismatches

Consider now examples like the following:

(29) A: No one in the department likes **you$_i$**!
      B: That's not true! Leslie does <like **me$_i$**>!

It seems that we have a mismatch here, since the pronominal element that must be in the ellipsis site is not identical to its correlate in the antecedent. In a similar way to how vehicle change phenomena are dealt with, the condition in Rudin 2019 captures cases like these by stipulating their acceptability in its definition of correlates in (10)b—the elements in bold are co-indexed and thus the pronoun in the ellipsis site finds a structure matching correlate.

---

[163] Other examples of vehicle change involve cases where a pronoun is exchanged in place of a *wh*-word:

(i)   a. The Deans told us who resigned, but they're not sure for what reasons <*who/they resigned>.
                                                                    (*Adapted from Chung et al. 2011*)
        b. That's a gazebo. But I don't know who built it or why <*who/they built it>.
                                                                    (*Adapted from Merchant 2001 201*)

A possible way to account for why the ellipsis sites in the above examples satisfy the identity condition proposed here is that *wh*-words are featurally a 3rd person pronominal plus a [WH] feature. Another case of vehicle change involve exchanging a name for a reflexive, as in the following:

(ii)   Monét likes Bob$_1$, and he$_1$ does too <like *Bob$_1$/himself$_1$>.

If the feature encoding reflexivity is privative, then it seems plausible that examples like (ii) would also satisfy the identity condition proposed here. In contrast to the judgments of speakers I have consulted, however, Fiengo & May 1994 and Oku 1998 claim that examples like (ii) are *ill*-formed (e.g., Oku 1998: 114, which provides '**Mary will admire John$_1$, and he$_1$ will <admire himself$_1$> too*'. Oku proposes an account to derive this kind of example and other related ones (see Oku 1998: 112-126). I leave for the future a thorough evaluation of these examples in relation to the proposal in the present work.

Once again, I bring this up to highlight how challenging data like the above are for identity conditions that propose a syntactic component.

What interests us is how the ellipsis site in (29)B satisfies the identity condition proposed in this dissertation. I suggest that these cases could be derived under my approach given a fine-grained featural decomposition of pronouns, as in Harley & Ritter (2002) and McGinnis (2005). As long as 3rd person is the absence of certain features, and 1st and 2nd person are differentiated only by the additional presence of a privative feature (i.e., [AUTHOR] for 1st as opposed to 2nd), then these cases might satisfy featural non-distinctness. Note, however, that these indexical switches need to be reconciled with our discussion of [PERSON] mismatches on T (see chapter 3). I leave a thorough analysis of these examples for future work, noting merely a way forward.

5.5 Strict identity—Warner's Effects

Warner 1986 and Lasnik 1995 discuss cases of mismatches involving non-identical forms of the English auxiliary verbs *be* and *have*. To illustrate the issue, observe that *be* appears to require strict identity, in contrast to lexical verbs like *sleep*:

(30)   *Warner's effects*
         a. Monét **slept** on stage, and Bob will <**sleep** on stage> too.
         b. *Monét **was** on stage, and Bob will <**be** on stage> too.

Lasnik 1995 proposes that facts like the above are explained if (i) strict identity is required in ellipsis and (ii) auxiliaries are pulled from the lexicon fully inflected, whereas other verbs attain their inflection derivationally (for alternatives to Lasnik's analysis, including empirical qualifications to the data in (30), see Potsdam 1996, Roberts 1998, and Lasnik 1999). At one point in the derivation, then, lexical verbs are indeed identical, whereas auxiliaries never are. If we take seriously the empirical picture that was painted in chapters

2-4, which I argued showed that strict identity is too strong a requirement, then the existence of Warner's Effects is a challenge to the proposed identity condition. In brief, we need to find an explanation for the ill-formed status of examples like (30)b that does not rely on a violation of the identity condition.

While I set aside a thorough discussion of how Warner's effects are to be accounted for, I do note an interesting observation in Merchant 2015 that suggests that Lasnik 1995's interpretation of the data might require some revision. Merchant shows that the ill-formed status of examples like (30) is maintained even under code-switching, where the antecedent of a bare form of the auxiliary *be* is in Greek:

(31)   *Warner's effects maintained under code-switching*
      a. Maria will be at the party, and her sister will (be), too.
      b. I   Maria **tha ine**   sto   party, and her sister **will** (**be**), too.
         the Maria **FUT be.NONPAST**.3S at.the party
      c. Maria is at the party, and her sister will *(be), too.   *Warner's Effect*
      d. I   Maria **ine**   sto   parti, and her sister **will** *(**be**), too.
         the Maria **be.NONPAST.3S** at.the party   *Warner's Effect under code-switching*

Setting aside that code-switching introduces its own set of complications, it seems that an account of Warner's Effects that is English specific and governed by whether English lexical items enter the derivation uninflected or not from the lexicon, is incomplete. Merchant sketches an account for facts like the above that attributes the ill-formed status of (31)c,d to a violation of a requirement that tense variables in the antecedent and target clause be bound from parallel positions. Much more needs to be said, of course, but I merely highlight that it is unclear whether Warner's Effects are an incontrovertible challenge to the identity condition proposed here.

5.6 Evidence for a subset condition?

Some recent work has argued for a one-way subset identity condition on ellipsis. This type of requirement is stricter than the one argued for in this dissertation, since it rules out configurations in which there are features in the ellipsis site that are not present in the antecedent. Proposals of this type would incorrectly rule out some of the data illustrated in this dissertation, if the interpretation of the facts argued for here is on the right track—recall the polarity mismatches in chapter 3, as well as the AF-Active mismatches in Kaqchikel in chapter 2 and certain [GENDER] mismatches in chapter 4 (see 5.2 above on Chung's generalization). For example, Murphy 2016 posits the following (see also my discussion of Donatelli 2019 in chapter 4):[164]

(32)   *Subset condition on ellipsis (Murphy 2016)*
        The morphosyntactic features in the ellipsis site must be a proper subset of those in the antecedent ($F_E \subset F_A$)

Data which has been discussed as support for such an approach include the following, from Rooryck and Schoorlemmer 2014 (cited by Murphy 2016):

(33)   a. Mary is leaving and I can see that Peter already has <left>.
        b. ?? Mary hasn't left yet, but I can see that Peter already is <leaving>.

The above shows an asymmetrical mismatch pattern—whereas a verb in the progressive form can antecede a participle, it seems that the inverse configuration is ill-formed.

---

[164] This condition rules out ellipses where the ellipsis site is syntactically *identical* to the antecedent. This is too restrictive, since ellipsis is clearly allowed in such case; a simple example like '*Gene saw two **movies** last night, while Roger saw three <**movies**>*' demonstrates this. Nevertheless, this is the approach defended by Murphy (see footnote 4 of Murphy 2016, where the author notes that his approach is different to Rooryck & Schoorlemmer' 2014 and Saab's 2015 proposals, which argue for a *subset* relation, not a *proper* subset one).

Rooryck and Schoorlemmer's interpretation of data like the above is that the progressive form has an additional [MOOD] feature—an example like (33)b thus violates a subset identity condition.

How robust are data like the above? Interestingly, it seems that the empirical picture is not nearly as clear cut as suggested in Murphy 2016. Thoms 2019 reports an identical configuration to (33)b that seems much more acceptable—in the example below, a progressive is elided and its antecedent is a participle (the example is attributed to David Pesetsky by way of Bronwyn Bjorkman):[165]

(34)   In the past, Mary hasn't **taken** medication for her condition, but she is <**taking** medication for her condition> now.

If we adopt Rooryck & Schoorlemmer's analysis of the progressive, the well-formed status of (34) seems to show that actually, an ellipsis site *can* contain a feature that is not present in the antecedent—as predicted by the identity condition proposed here.

We have an empirical conflict, then—faced with (33)b, the identity condition proposed in this dissertation appears to overgenerate. In contrast, a stricter subset condition appears to undergenerate, since it incorrectly rules out (34). While I will not commit to an account for the ill-formed status of (33)b, it seems plausible to attribute its deviancy to an issue that

---

[165] It seems that judgements where there is a progressive in the ellipsis site seem to be variable in general. As Thoms 2019 discusses, example (i) is judged unacceptable in Lasnik 1995, but some speakers do accept it, based primarily on whether the remnant auxiliary is stressed or not. Furthermore, Thoms gives (ii) as acceptable. Observe that there is a verb in its progressive form in the ellipsis site:

(i) *John won't enter the competition, but Peter is <entering the competition>.
                                        (Lasnik 1995; attributed to Quirk et al. 1972)
(ii) Emma intends to write a novel, and Anna already is <writing a novel>.

is orthogonal to the identity condition—in fact, the acceptability of (34) might force us to pursue such an analysis.

Moving on, Murphy 2016 also gives the following examples illustrating a categorial mismatch as evidence for the subset condition:

(35)   a. *Eddy has already **reported** on the accident, but we cannot find Tim's <**report on the accident**>.
        b. *We read Tim's **report** on the accident, but Eddy hasn't <**reported** on the accident>.

The data above seem to suggest that NP ellipsis is ill-formed when the nominal in the ellipsis site has a verbal antecedent (35)a; conversely, VP ellipsis is ill-formed when the verb in the ellipsis site has a nominal antecedent (35)b. A subset condition on ellipsis is violated here, since there is no subset relation between the content of the ellipsis sites and the antecedent.

These data, however, cannot adjudicate between a subset condition and the proposal in this dissertation. A plausible analysis of the above facts is that there is a clash in categorial heads—the nominal and verbal forms that are mismatching differ on whether the single underlying √ROOT was categorized by a verbalizing *v* or a nominalizing *n*. Under my approach, it is the clash of these heads that leads to a violation of featural non-distinctness, correctly ruling out the data. [166]

---

[166] Omer Preminger points out that there are cases of categorial mismatches that are ill-formed and seem to be problematic for a subset condition and the condition proposed in this dissertation. Consider how the nominal *nationalization* and the verb *nationalize* cannot mismatch under ellipsis. Unlike in the examples in (35), it seems plausible that *nationalization* is derived by adding a nominalizing head on top of the [$_{vP}$ [v √NATIONALIZE]] structure, but an example like the following is nevertheless ill-formed:

The empirical picture regarding the interaction of nominal and verbal forms under ellipsis is complicated, however, by examples that seem to show that mismatches of a similar flavor are actually well-formed in some instances. Observe the data below (see also Johnson 2001 for more examples):

(36)   A: Do you know who the forger is?
         B: I talked to Seth Frank. Apparently Whitney learned how in prison.

Merchant 2001: 22

(37)   a. ??That man is a robber, and when he does <rob> he tries not to make any noise.
         b. *That man is a thief, and when he does <steal>, he tries not to make any noise.

Merchant 2013c

Examples like the above seem to be challenging for any syntactic approach to the identity condition.

However, it seems puzzling that the particular lexical items that are manipulated in examples of this nature seems to matter for the ultimate well- or ill-formed status of the examples. For instance, if we use the alternation *teacher ~ to teach*, an example akin to (36) is not acceptable to my consultants; similarly, if we use *driver ~ drive* for an example similar to (37), the result is similarly much degraded:

(38)   A: Do you know who the teacher is?
         B: *I Asked Peggy. She says Sig learned how <to teach> at Concordia.

(39)   *That man is indeed a driver, but when he does <drive>, he scares others on the road.

If the particular lexical items that are manipulated seems to influence the acceptability of examples of this kind, then it seems that the apparent well-formed status of examples in (36)-(37) deserves closer scrutiny.

(i) *My party opposed the nationalization of resources, but the ruling coalition went ahead and did, anyway. It seems that categorial mismatches are a ripe area for future study.

5.7 Summary

In this chapter, I discussed areas for future research, suggesting how the proposal in this dissertation could be expanded beyond the three empirical domains that were discussed in chapters 2-4. I hope that the suggestions here entice further work into the potential applicability of my proposal to shed light on a range of puzzles that could ultimately adjudicate between different formulations of the identity condition on ellipsis.

Chapter 6: Summary of results

At the outset of this dissertation, the following question was raised:

(1) *The identity question in this dissertation*
   What is the syntactic relationship between the understood material in the ellipsis site and its antecedent?

On the basis of evidence from a variety of languages and elliptical constructions, I arrived at the following answer:

(2) *Syntactic identity condition on ellipsis (final version)*
   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.

In chapter 2, I introduced novel sluicing data from Kaqchikel to the identity debate, demonstrating that a subset of voice mismatches are well-formed in the language—in stark contrast to languages like English. I proposed that all well-formed mismatches involve an antecedent or ellipsis site lacking VoiceP—the so-called Agent Focus voice—thus satisfying (2)a. In chapter 3, I discussed an asymmetry in mismatch availability in English that appears (at first glance) to be dependent on the locus of the mismatching element: elements within the eventive core (VoiceP/vP) cannot mismatch, whereas elements above it can. I argued that proposals that seek to divide the clause into areas that enter into the calculation of identity in ellipsis and areas that do not are untenable. Instead, the type of mismatch matters, and (2)a can derive the abovementioned asymmetry, plus a range of cross-linguistic data showing that featural clashes are banned regardless of the locus of the mismatching feature. Chapter 4 discussed [GENDER] and [NUMBER] mismatches in the nominal domain. I showed that an intricate tripartite asymmetry in mismatch behavior across a range of languages could be derived by (2)a-b, when coupled with (i) a

decompositional view of nominals where [GENDER] is on *n*, (ii) repair-by-ellipsis of morphophonological gaps, and (iii) the impossibility of repair-by-ellipsis of grammatical violations unrelated to externalization. Finally, chapter 5 discussed possible extensions and potential challenges for the identity condition in (2): (i) "voice" mismatches in Chamorro and Malagasy sluicing, (ii) Chung's generalization, (iii) vehicle change phenomena, (iv) indexical switches, (v) Warner's effects, and (vi) a range of data that seems at first glance to suggest that, in contrast to (2), a subset condition on ellipsis is warranted instead. I argued that the proposal advanced here holds promise in providing an account of these empirical domains as well, and it is my hope that the present work provides a starting point for yet further explorations into the syntactic component of the identity condition.

References

Abe, Jun. 2015. *The In-Situ Approach to Slucing*. Linguistik Aktuell – Linguistics Today 222. Amsterdam: John Benjamins.

Abels, Klaus. 2003. *Successive Cyclicity, Anti-locality, and Adposition Stranding.* PhD Dissertation, University of Connecticut at Storrs.

Abels, Klaus. 2019a. "Movement and islands". In Jeroen van Craenenbroeck and Tanja Temmerman.(eds.) *The Oxford Handbook of Ellipsis*. Oxford & New York: Oxford University Press.

Abels, Klaus. 2019b. "On "sluicing" with apparent massive pied-piping". *Natural Language and Linguistic Theory* 37: 1205-1271.

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

Adams, Pern Wang, and Satoshi Tomioka. 2012. "Sluicing in Mandarin Chinese: An Instance of Pseudo-Sluicing." In Jason Merchant & Andrew Simpson (eds.), *Sluicing: Cross-Linguistic Perspectives*. Oxford: Oxford University Press.

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

Aissen, Judith L. 1992. "Topic and Focus in Mayan." *Language* 68 (1): 43–80.

Aissen, Judith. 1996. "Pied-piping, abstract agreement, and functional projections in Tzotzil". *Natural Language and Linguistic Theory* 14: 447-491.

Aissen, Judith. 1999. "Agent Focus and Inverse in Tzotzil". *Language* 75: 451-485.

Aissen, Judith. 2011. "On the Syntax of Agent Focus in K'ichee'." In Kirill Shklovsky, Pedro Mateo Pedro, and Jessica Coon (eds.), *Proceedings of Formal Approaches to*

Mayan Linguistics I. MIT Working Papers in Linguistics Vol. 63. Cambridge, MA: MITWPL. 1-16.

Aissen, Judith. 2017a. "Information Structure in Mayan". In Judith Aissen, Nora England, & Roberto Zavala (eds.), *The Mayan Languages*. London: Routledge.

Aissen, Judith. 2017b. "Correlates of Ergativity in Mayan." In Jessica Coon, Diane Massam, & Lisa Travis (eds.), *The Oxford Handbook of Ergativity*. Oxford: Oxford University Press.

Aissen, Judith, Nora England, & Roberto Zavala, eds. 2017. *The Mayan Languages*. London: Routledge.

Aldridge, Edith. 2004. *Ergativity and Verb-Initial Word Order in Austronesian Languages*. PhD Dissertation, Cornell University.

Alexiadou, Artemis. 2004. "Inflection class, gender, and DP internal structure". In Gereon Müller, Lutz Gunkel, and Gisela Zifonum (eds.), *Explorations in Nominal Inflection*. Berlin: Mouton.

Alexiadou, Artemis. 2017. "Gender and nominal ellipsis". In Nicholas LaCara, Keir Moulton & Anne-Michelle Tessier (eds.), *A Schrift to Fest Kyle Johnson.* 11-22.

Anagnostopoulou, Elena. 2003. *The Syntax of Ditransitives: Evidence from Clitics*. Studies in Generative Grammar 54. Berlin-New York: Mouton de Gruyter.

Anand, Pranav, Daniel Hardt & James McCloskey. 2021. *The Domain of Matching in Sluicing*. Ms.Aavailable at: https://ling.auf.net/lingbuzz/005674

Anderbois, Scott. 2008. *Sluicing and the Nature of the Antipassive in Yukatek Maya*. Ms. University of California, Santa Cruz.

Arad, Maya. 2005. *Roots and patterns: Hebrew morpho-syntax*. Dordrecht: Kluwer.

Aronoff, Mark. 1994. *Morphology by Itself: Stems and Inflectional Classes.* Cambridge, MA: MIT Press.

Assmann, Anke, Doreen Georgi, Fabian Heck, Gereon Müller, & Philipp Weisser. 2015. "Ergatives Move Too Early: On an Instance of Opacity in Syntax." *Syntax* 18 (4): 343–87.

Baier, Nicholas Benson. 2018. *Anti-Agreement*. PhD Dissertation, University of California at Berkeley.

Baltin, Mark. 2012. "Deletion versus pro forms: An overly simple dichotomy?". *Natural Language and Linguistic Theory* 30: 381-423.

Banerjee, Neil. 2020. "Ellipsis as obliteration: Evidence from Bengali negative allomorphy". In *Proceedings of the Linguistic Society of America* 5(1): 133-143.

Barrie, Michael. 2016. "Gender Mismatches and Ellipsis in Cayuga". *Journal of Cognitive Science* 17-3: 361-388.

Barros, Matt. 2014. *Sluicing and Identity in Ellipsis*. PhD Dissertation, Rutgers University.

Béjar, Susana, and Milan Rezac. 2003. "Person Licensing and the Derivation of PCC Effects." In *Amsterdam Studies in the Theory and History of Linguistic Science Series 4*, 49–62.

Bennett, Ryan. 2016. "Mayan Phonology." *Language and Linguistics Compass* 10 (10): 469–514.

Bennett, Ryan, Emily Elfner & James McCloskey. 2019. "Prosody, focus, and ellipsis in Irish". *Language* 95(1): 66-106.

Benveniste, Emile. 1971. "The nature of pronouns". In *Problems in general linguistics*, Coral Gables, FL: University of Miami Press. 217–22.

Bermúdez-Otero, Ricardo. 2013. "The Spanish lexicon stores stems with theme vowels, not roots with inflectional class features". *Probus* 25: 3-103.

Bobaljik, Jonathan David. 2012. *Universals in Comparative Morphology: Suppletion, Superlatives, and the Structure of Words*. Cambridge, MA: MIT Press.

Bobaljik, Jonathan & Cynthia Levart Zocca. 2011. "Gender markedness: the anatomy of a counterexample". *Morphology* 21: 141-166.

Boeckx, Cedric & Howard Lasnik. 2006. "Intervention and Repair". *Linguistic Inquiry* 37 (1): 150-155.

Borer, Hagit. 1984. *Parametric Syntax: Case Studies in Semitic and Romance Languages*. Dordrecht: Foris.

Borer, Hagit. 2005. *Structuring Sense Volume I: In Name Only*. Oxford: Oxford University Press.

Borer, Hagit. 2014. "Wherefore roots?" *Theoretical Linguistics* 40(3/4): 343-359.

Bresnan, Joan. 1972. "Theory of Complementation in English Syntax." PhD dissertation, MIT.

Brucart, José María. 1987. *La Elisión Sintáctica En Español*. Barcelona: Bellaterra.

Brucart, José María. 1999. "La elipsis". In Ignacio Bosque & Violeta Demonte (eds.), *Gramática descriptiva de la lengua española*. Madrid: Espasa-Calpe. 2787-2863.

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*. Blackwell Publishing. 579-601.

Burukina, Irina. 2019. "Reflexive Functional Head, Verbal and Nominal Predicates."
In *Proceedings of the 36ᵗʰ West Coast Conference on Formal Linguistics*. Somerville,
MA: Cascadilla Proceedings Project.

Can Pixabaj, Telma & Nora C. England. 2011. "Nominal Topic and Focus in K'ichee'."
In Rodrigo Gutiérrez-Bravo, Line Mikkelsen, and Eric Potsdam (eds.), *Representing
Language: Essays in Honor of Judith Aissen*. Santa Cruz: Linguistics Research Center,
UC Santa Cruz. 15–30

Carstens, Vicki. 2016. "Delayed Valuation: a reanalysis of "upwards" complementizer
agreement and locality". *Syntax* (19): 1. 1-42.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Chomsky, Noam. 1972. "Some Empirical Issues in the Theory of Transformational
Grammar." In Paul Stanley Peters (ed.), *Goals of Linguistic Theory*. Englewood Cliffs:
Prentice Hall Inc.

Chomsky, Noam. 2001. "Derivation by Phase." In Michael Kenstowicz (ed.), *Ken Hale: A
Life in Language*. Cambridge, MA: MIT Press. 1–50.

Chomsky, Noam. 2008. "On Phases." In Robert Freidin, Carlos Otero, & Maria Luisa
Zubizarreta (eds.), *Foundational Issues in Linguistic Theory*. Cambridge, MA: MIT
Press. 133–66.

Chung, Sandra. 2006. "Sluicing and the Lexicon: The Point of No Return." In R. Cover
and Y. Kim (eds.), *Proceedings of BLS 31*. 73–91.

Chung, Sandra. 2013a. "Syntactic Identity in Sluicing: How Much and Why." *Linguistic
Inquiry* 44 (1): 1–44.

Chung, Sandra. 2013b. "The Syntactic Relations behind Agreement". In Lisa Lai-Shen Cheng & Norbert Corver (eds.), *Diagnosing Syntax*. Oxford: Oxford University Press.

Chung, Sandra, William Ladusaw, and McCloskey Jim. 1995. "Sluicing and Logical Form." *Natural Language Semantics* 3 (3): 239–82.

Chung, Sandra, William Ladusaw, and McCloskey Jim. 2011. "Sluicing(:) between structure and inference". In Rodrigo Gutiérrez-Bravo, Line Mikkelsen, & Eric Potsdam (eds.), *Representing Language: Essays in Honor of Judith Aissen*. 31-50.

Cinque, Guglielmo. 1999. *Adverbs and Functional Heads: A Cross-Linguistic Perspective*. Oxford and New York: Oxford University Press.

Citko, Barbara. 2014. *Phase Theory: An Introduction*. Cambridge: Cambridge University Press.

Clemens, Lauren, and Jessica Coon. 2018. "Deriving Verb-Initial Word Order in Mayan." *Language* 94 (2): 237–80.

Collins, Chris. 2018. "The Logic of Contextual Allomorphy". Ms., NYU. Available at https://ling.auf.net/lingbuzz/004176.

Conner, Tracy. 2019. *Divorce Licensing: Separate Criteria for Predicate and Clausal Ellipsis*. PhD Dissertation, University of Massachusetts at Amherst.

Coon, Jessica. 2016. "Mayan Morphosyntax." *Language and Linguistics Compass* 10 (10): 515–50.

Coon, Jessica. 2017. "Little-v Agreement and Templatic Morphology in Ch'ol." *Syntax* 20: 101–37.

Coon, Jessica, Pedro Mateo Pedro, and Omer Preminger. 2014. "The Role of Case in A-Bar Extraction Asymmetries." *Linguistic Variation* 14 (2): 179–242.

Coon, Jessica, Nico Baier & Theodore Levin. 2021. "Mayan Agent Focus and the Ergative Extraction Constraint: Facts and Fictions Revisited". *Language* 97(2): 269-332.

Conrod, Kirby. 2019. *Pronouns raising and emerging*. PhD Dissertation, University of Washington.

Corbett, Greville. 1991. *Gender*. Cambridge: Cambridge University Press.

van Craenenbroeck, Jeroen. 2010a. "Invisible Last Resort: A note on clefts as the underlying source for sluicing". *Lingua* 120(7): 1714-1726.

van Craenenbroeck, Jeroen. 2010b. *The syntax of ellipsis: Evidence from Dutch dialects*. New York: Oxford University Press.

van Craenenbroeck, Jeroen. 2012. "How Do You Sluice When There Is More than One CP?" In Jason Merchant & Andrew Simpson (eds.), *Sluicing:Cross-Linguistic Perspectives*. Oxford: Oxford University Press. 40–67.

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.

Craig, Collette Grinevald. 1979. "The antipassive and Jacaltec". In Laura Martin (ed.), *Papers in Mayan linguistics*, vol. 1. Lucas Brothers. 139–164

Creemers, Ava, Jan Don, & Paula Fenger. 2018. "Some affixes are roots, others are heads". *Natural Language and Linguistic Theory* 36: 45-84.

Culicover, Peter W. & Ray Jackendoff. 2019. "Ellipsis in Simpler Syntax". In Jeroen van Craenenbroeck & Tanja Temmerman (eds.), *The Oxford Handbook of Ellipsis*. Oxford: Oxford University Press.

D'Alessandro, Roberta. 2021. "A short history of Agree". To appear in *The Cambridge Handbook of Minimalism*. Available at: https://ling.auf.net/lingbuzz/005888

Dayley, Jon P. 1981. "Voice and Ergativity in Mayan Languages". *Journal of Mayan Linguistics* 2: 6-82.

Dayley, Jon P. 1985. *Tzutujil Grammar*. Berkeley: University of California Press.

Davies, William D. & Luis Enrique Sam-Colop. 1990. "K'iche' and the structure of antipassive". *International Journal of American Linguistics* 66: 522–549.

Deacon, Robert Joel. 2014. *Adpositions in Distributed Morphology: The Nature of √ROOTS and Categorial Heads.* PhD Dissertation, University of Florida.

Deacon, Robert Joel. 2017. "Roots in adpositional domains: Reasons to include a little-p head". In Proceedings of the *34th West Coast Conference on Formal Linguistics*. Somerville, MA: Cascadilla Proceedings Project. 131-140.

den Dikken, Marcel. 2020. *Unmatched and Unparalleled: Voice and argument structure mismatches in ellipsis—Analysis and implications*. Ms., Eötvös Loránd University & Research Institute for Linguistics.

Depiante, Marcela. 2000. *The syntax of deep and surface anaphora: A study of null complement anaphora and stripping/bare argument ellipsis*. PhD Dissertation, University of Connecticut at Storrs.

Depiante, Marcela. 2004. "Dos casos de elipsis con partícula de polaridad en español". *RASAL* 1: 53–69.

Depiante, Marcela. 2019. "Null Complement Anaphora". In Jeroen van Craenenbroeck & Tanja Temmerman (eds.), *The Oxford Handbook of Ellipsis*. Oxford: Oxford University Press.

Depiante, Marcela & José Masullo. 2001. "Género y número en la elipsis nominal: consecuencias para la hipótesis lexicalista". Paper presented at the *1st Encuentro de Gramática Generativa.*

Donatelli, Lucia Elizabeth. 2019. *The morphosemantics of Spanish gender: Evidence from small nominals*. PhD Dissertation, Georgetown University.

Douglas, Jamie, Rodrigo Ranero, and Michelle Sheehan. 2017. "Two Kinds of Syntactic Ergativity in Mayan." In Michael Yoshitaka Erlewine (ed.), *Proceedings of GLOW in Asia XI, Volume 2*. Cambridge MA: MIT Working Papers in Linguistics. 41–56

Drummond, Emily. 2021. "Maintaining syntactic identity under sluicing: Pseudoclefts and voice (mis)matches". To appear in *Proceedings of the 39th West Coast Conference on Formal Linguistics*.

Elbourne, Paul. 2005. *Situations and Individuals*. Cambridge, MA: MIT Press

Embick, David. 2010. *Localism versus Globalism in Morphology and Phonology*. Cambridge, MA: MIT Press.

Embick, David. 2015. *The Morpheme: A Theoretical Introductions*. Interface Explorations. Berlin: Mouton de Gruyter.

Embick, David & Rolf Noyer. 2007. "Distributed morphology and the syntax/morphology interface". In Gillian Ramchand & Charles Reiss (eds.), *The Oxford Handbook of Linguistic Interfaces*.Oxford: Oxford University Press.

England, Nora. 1991. "Changes in Basic Word Order in Mayan Languages." *International Journal of American Linguistics* 57 (4): 446–86.

Erlewine, Michael Yoshitaka. 2016. "Anti-Locality and Optimality in Kaqchikel Agent Focus." *Natural Language and Linguistic Theory* 34 (2): 429–79.

Erlewine, Michael Yoshitaka & Theodore Levin. 2017. "On the unavailability of argument ellipsis in Kaqchikel". To appear in *Proceedings of FAMLi 4.*

Erlewine, Michael Yoshitaka, Theodore Levin & Coppe van Urk. 2017. "Ergativity and Austronesian-Type Voice Systems". In Jessica Coon, Diane Massam, & Lisa Travis (eds.), *The Oxford Handbook of Ergativity*. Oxford: Oxford University Press.

Erschler, David. 2018. *Typology of Bizarre Ellipsis Varieties*. PhD Dissertation, University of Massachusetts Amherst.

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

Fortin, Catherine. 2019. "Indonesian". In Jeroen van Craenenbroeck & Tanja Temmerman (eds.), *The Oxford Handbook of Ellipsis*. Oxford: Oxford University Press.

Fox, Danny & Howard Lasnik. 2003. "Successive Cyclic Movement and Island Repair: The Difference between Sluicing and VP-Ellipsis." *Linguistic Inquiry* 34: 143–154.

Frampton, Jon & Sam Gutmann. 2000. *Agreement is Feature Sharing*. Ms. Northeastern University.

Frampton, Jon & Sam Gutmann. 2006. "How sentences grow in the mind: Agreement and selection in an efficient minimalist syntax". In Cedric Boeckx (ed.), *Agreement Systems*. Amsterdam: John Benjamins, 121-57.

Gallego, Ángel. 2014. "Deriving Feature Inheritance from the Copy Theory of Movement". *The Linguistic Review* 31: 41-71.

García Álvarez, Iván. 2008. *Generality and exception: a study in the semantics of exceptives*. PhD Dissertation, Stanford University.

García Matzar, Lolmay Pedro. 2007. *Derivación de Palabras En Kaqchikel*. Guatemala: Editorial Cholsamaj.

García Matzar, Lolmay Pedro & José Domingo Rodríguez Guaján. 1997. *Rukemik ri Kaqchikel Chi': Gramática Kaqchikel*. Guatemala: Editorial Cholsamaj.

Goldberg, Adele E. & Florent Perek. 2019. "Ellipsis in Construction Grammar". In Jeroen van Craenenbroeck & Tanja Temmerman (eds.), *The Oxford Handbook of Ellipsis*. Oxford: Oxford University Press.

Grebenyova, Lydia. 2006. *Multiple Interrogatives: Syntax, Semantics, and Learnability*. PhD Dissertation, University of Maryland at College Park.

Gribanova, Vera. 2013. "Verb-stranding verb phrase ellipsis and the structure of the Russian verbal complex". *Natural Language and Linguistic Theory* 31: 91-136.

Gribanova, Vera. 2018. *Head movement, ellipsis, and identity*. Ms., Stanford University.

Grinder, John & Paul Postal. 1971. "Missing Antecedents". *Linguistic Inquiry* 2(3): 269-312.

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*. Cambridge, MA: MIT Press. 111–76.

Halle, Morris & Alec Marantz. 1994. "Some key features of Distributed Morphology". *MIT Working Papers in Linguistics* 21: 275-288.

Hankamer, Jorge & Ivan Sag. 1976. Deep and Surface Anaphora. *Linguistic Inquiry* 7(3): 391-428.

Hankamer, Jorge. 1978. "On the Nontransformational Derivation of some Null VP Anaphors". *Linguistic Inquiry* 9 (1): 66-74.

Harley, Heidi. 2014. "On the Identity of Roots." *Theoretical Linguistics* 40 (3): 225–76.

Harley, Heidi. 2017. "The 'Bundling' Hypothesis and the Disparate Functions of Little v." In Roberta D'alessandro, Irene Franco, & Angel J. Gallego (eds.), *The Verbal Domain*. Oxford: Oxford University Press.

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

Harris, James. 1998. "Spanish imperatives: syntax meets morphology". *Journal of Linguistics* 34: 27-52.

Harris, Zelig. 1982. *A Grammar of English on Mathematical Principles*. NY: Jon Wiley & Sons.

Haspelmath, Martin. 2006. "Against markedness (and what to replace it with)". *Linguistics* 42(1): 25-70.

Heaton, Raina. 2017. *A Typology of Antipassives, with Special Reference to Mayan*. PhD Dissertation, University of Hawaii at Manoa.

Heaton, Raina, Kamil Deen, and William O'Grady. 2016. "An Investigation of Relativization in Kaqchikel Maya." *Lingua* 170: 35–46.

Henderson, Robert & Jessica Coon. 2018. "Adverbs and Variability in Kaqchikel Agent Focus: A Reply to Erlewine (2016)." *Natural Language and Linguistic Theory* 36 (1): 149–73.

Hirata, Ryuchiro. 2021. *Preposition Stranding in Welsh*. PhD Dissertation, Bangor University.

Hoeksema, Jack. 1987. "The logic of exception". In A. Miller & J. Powers (eds.), Proceedings of ESCOL 4. Columbus, OH: The Ohio State University. 100-113.

Holmberg, Anders & Thorbjörg Hróasdóttir. 2003. "Agreement and movement in Icelandic raising constructions". *Lingua* 113: 997-1019.

Hornstein, Norbert, Howard Lasnik & Juan Uriagereka. 2007. "The Dynamics of Islands: Speculations on the Locality of Movement". *Linguistic Analysis* 33 (1-2). 149-175.

Ingria, Robert JP. 1990. "The limits of unification". In *Proceedings of the 28th Annual Meeting of the Association for Computational Linguistics*.

Isac, Daniela. 2015. *The Morphosyntax of Imperatives*. Oxford & New York: Oxford University Press.

Jacobson, Pauline. 2016. "The short answer: Implications for direct compositionality (and vice-versa)". *Language* 92 (2). 331-375.

Johnson, Kyle. 2001. "What VP ellipsis can do, and what it can't, but not why". In Chris Collins & Mark Baltin (eds.), *The Handbook of Contemporary Syntactic Theory*. Malden, MA: Blackwell Publishing.

Johnson, Kyle. 2019. "Gapping and Stripping". In Jeroen van Craenenbroeck and Tanja Temmerman.(eds.) *The Oxford Handbook of Ellipsis*. Oxford & New York: Oxford University Press.

Katz, Jerrold J. & Paul J. Postal. 1964. *An integrated theory of linguistic descriptions*. Cambridge, MA: MIT Press.

Kehler, Andrew. 2002. *Coherence, reference, and the theory of grammar*. Center for the Study of Language and Information. Stanford, CA.

Kehler, Andrew. 2019. "Ellipsis and discourse". In Jeroen van Craenenbroeck and Tanja Temmerman (eds.) *The Oxford Handbook of Ellipsis*. Oxford & New York: Oxford University Press.

Kennedy, Chris & Jason Merchant. 2000. "Attributive comparative deletion". *Natural Language and Linguistic Theory* 18 (1): 89-146.

Kenstowicz, Michael. 2013. "Realize morpheme in Kaqchikel". In Michael Kenstowicz (ed.), *Studies in Kaqchikel Grammar*.

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

Kertz, Laura. 2013. "Verb Phrase Ellipsis: The View from Information Structure." *Language* 89 (3): 290–428.

King, Gareth. 2003. *Modern Welsh: A Comprehensive Grammar*. London: Routledge.

Kiparsky, Paul, Carol Kiparsky, D. D. Steinberg, & L. A. Jakobovits. 1971. "Fact." In *Semantics: An Interdisciplinary Reader in Philosophy, Linguistics, and Psychology*. Cambridge: Cambridge University Press. 345-369.

Konnelly, Lex., & Elizabeth Cowper. 2020. "Gender diversity and morphosyntax: An account of singular *they*". *Glossa: A Journal of General Linguistics*, *5*(1), 40. DOI: http://doi.org/10.5334/gjgl.1000

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.

Kratzer, Angelika. 1996. "Severing the external argument from its verb". In Johan Rooryck and Laurie Zaring (eds)., *Phrase Structure and the Lexicon*. Studies in Natural Language and Linguistic Theory 33. Dordrecht: Springer. 109-137.

Kroll, Margaret. 2018. "Polarity reversals under sluicing". In Robert Truswell, Chris Cummins, Caroline Heycock, Brian Rabern and Hannah Rohde (eds.), *Proceedings of Sinn und Bedeutung 21*.

Kroll, Margaret. 2019. "Polarity reversals under sluicing". *Semantics and Pragmatics* 12(18): 1-49.

Kučerová, Ivona. 2018. "φ-Features at the Syntax-Sematics Interface: Evidence from Nominal Inflection". *Linguistic Inquiry* 49(4): 813-845.

Laka, Itziar. 1990. *Negation in syntax: on the nature of functional categories and projections*. PhD Dissertation, MIT.

Laka, Itziar. 1991. "Negation in Syntax: On the Nature of Functional Categories and Projections". *International Journal of Basque Linguistics and Philology (ASJU)* XXV-1: 65-136.

Lakoff, George. 1972. "The Arbitrary Basis of Transformational Grammar." *Language* 48 (1): 76–87.

Landau, Idan. 2016. "DP-internal semantic agreement: A configurational analysis". *Natural Language and Linguistic Theory* 34(3): 975-1020.

Langacker, Ronald W. 1974. "Movement Rules in Functional Perspective". *Language* 50(4): 630-64.

Larsen, Tomas W. & William M. Norman. 1979. "Correlates of ergativity in Mayan grammar". In Frans Plank (ed.), *Ergativity: Towards a theory of grammatical relations*. London/ New York: Academic Press. 347-370.

Larson, Brooke & Norbert Hornstein. 2013. "A Note on P-Stranding and Adjunct Extraction from Nominals". *Linguistic Inquiry* 44 (4). 669-674.

Lasnik, Howard. 1988. "Subjects and the Theta-Criterion". *Natural Language and Linguistic Theory* 6: 1-17.

Lasnik, Howard. 1995. "Verbal morphology: syntactic structures meets the minimalist program". In Héctor Campos & Paula Kempchinsky (eds.), *Evolution and Revolution in linguistic Theory: Essays in Honor of Carlos Otero*. Georgetown University Press: Georgetown. 251-268.

Lasnik, Howard. 1999. "A Gap in an Ellipsis Paradigm: Some Theoretical Implications". *Linguistic Analysis* 27(3-4). 166-85.

Lasnik, Howard. 2001. "When Can You Save a Structure by Destroying It?" In Minjoo Kim & Uri Strauss (eds.), *Proceedings of the North East Linguistic Society* 31. GLSA. 301-20.

Lasnik, Howard. 2005. "How to Evade Moving Violations." Handout from a lecture at LSA Summer Institute, MIT, Cambridge, MA.

Lasnik, Howard. 2008. "On the Development of Case Theory: Triumphs and Challenges." In Robert Freidin, Carlos Otero, and María Luisa Zubizarreta (eds.), *Foundational Issues in Linguistic Theory*. Cambridge, MA: MIT Press. 17-41.

Lasnik, Howard. 2009. "Island Repair, Non-Repair and the Organization of the Grammar." In Kleanthes Grohmann (ed.), *InterPhases: Phase-Theoretic Investigations of Linguistic Interfaces*. Oxford: Oxford University Press. 339-53.

Lasnik, Howard, and Kenshi Funakoshi. 2019. "Ellipsis in Transformational Grammar." In Jeroen van Craenenbroeck and Tanja Temmerman (eds.), *The Oxford Handbook of Ellipsis*. Oxford: Oxford University Press.

Legate, Julie Anne. 2008. "Morphological and Abstract Case." *Linguistic Inquiry* 39 (1): 55–101.

Legate, Julie Anne. 2014. *Voice and v: Lessons from Acehnese*. Cambridge, MA: MIT Press.

Levin, Lori S. 1982. "Sluicing: A lexical interpretation procedure". In Joan Bresnan (ed.), *The mental representation of grammatical relations*. Cambridge, MA: MIT Press. 590-654.

Lipták, Anikó. 2013. "Identity in ellipsis: An introduction". *Lingua* 166: 155-171.

Lobeck, Anne. 1995. *Ellipsis: Functional Heads, Licensing, and Identification*. Oxford: Oxford University Press.

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.

Lyskawa, Paulina. 2021. *Coordination without grammar-internal feature resolution*. PhD Dissertation, University of Maryland at College Park.

Marantz, Alec. 1995. *Cat as a phrasal idiom: Consequences of late insertion in Distributed Morphology*. Ms., MIT.

Marantz, Alec. 1997. "No Escape from Syntax: Don't Try Morphological Analysis in the Privacy of Your Own Lexicon". *U. Penn Working Papers in Linguistics* Volume 4.2: 201-25.

Martin, Roger. 2001. "Null case and the distribution of PRO". *Linguistic Inquiry* 32(1): 141-66.

Masullo, Pascual José & Marcela Depiante. 2004. "Gender is in the Lexicon, Number is in the Syntax: Evidence from Nominal Ellipsis in Spanish". Handout of a talk presented at *GLOW 2004*, Aristotle University, Thessaloniki.

Mathieu Éric, Myriam Dali, & Gita Zareikar (eds.). 2018. *Gender and Noun Classification*. Oxford: Oxford University Press.

McGinnis, Martha. 2005. "On markedness asymmetries in person and number". *Language* 81(3): 699-718.

McKenna Brown, R., Judith M. Maxwell, and Walter Little. 2006. *¿La Ütz Awäch? Introduction to the Kaqchikel Maya Language.* Austin: University of Texas Press.

Mendes, Gesoel. 2020. *Investigations on Salvation and Non-Salvation by Deletion*. PhD Dissertation, University of Maryland at College Park.

Mendes, Gesoel & Jason Kandybowicz. 2021. "Salvation by Deletion in Nupe". To appear in *Linguistic Inquiry*.

Mendes, Gesoel & Andrew Nevins. 2021. "When ellipsis can save defectiveness and when it can't". To appear in *Linguistic Inquiry*.

Mendes, Gesoel & Rodrigo Ranero. 2021. "Chain Reduction via Substitution: Evidence from Mayan." *Glossa: A Journal of General Linguistics*. 6 (1), 10. DOI: http://doi.org/10.5334/gjgl.1087

Merchant, Jason. 1998. "Pseudosluicing: Elliptical Clefts in Japanese and English." In Artemis Alexiadou, Nanna Fuhrhop, Paul Law, & Ursula Kleinhenz (eds.), *ZAS Papers in Linguistics 10*. Zentrum fuer Allgemeine Sprachwissenschaft, Berlin. 88–112.

Merchant, Jason. 2001. *The Syntax of Silence: Sluicing, Islands and the Theory of Ellipsis*. Oxford: Oxford University Press.

Merchant, Jason. 2002. "Swiping in Germanic." In Jan-Wouter Zwart & Werner Abraham (eds.), *Studies in Comparative Germanic Syntax*. Amsterdam: John Benjamins. 295–321.

Merchant, Jason. 2005. "Revisiting syntactic identity conditions". Talk given at University of California at Berkeley, workshop on ellipsis, October 2005.

Merchant, Jason. 2008. "Variable Island Repair under Ellipsis." In Kyle Johnson (ed.), *Topics in Ellipsis*. Cambridge: Cambridge University Press. 132–53.

Merchant, Jason. 2013a. "Voice and Ellipsis." *Linguistic Inquiry* 44 (1): 77–108.

Merchant, Jason. 2013b. "Diagnosing ellipsis". In Lisa Lai-Shen Cheng & Norbert Corver (eds.), *Diagnosing Syntax*. Oxford: Oxford University Press.

Merchant, Jason. 2013c. "Polarity items under ellipsis". In Lisa Lai-Shen Cheng & Norbert Corver (eds.), *Diagnosing Syntax*. Oxford: Oxford University Press.

Merchant, Jason. 2014. "Gender mismatches under nominal ellipsis". *Lingua* 151: 9-32.

Merchant, Jason. 2015. "On ineffable predicates: Bilingual Greek-English code-switching under ellipsis". *Lingua* 166: 199-213.

Merchant, Jason. 2019. "Ellipsis: A Survey of Analytical Approaches". In Jeroen van Craenenbroeck & Tanja Temmerman (eds.), *The Oxford Handbook of Ellipsis*. Oxford: Oxford University Press.

Mitrović, Moreno & Phoevos Panagiotidis. 2020. "Adjectives exist, adjectivisers do not: a bicategorial typology". Glossa: A Journal of General Linguistics, 5(1), 58.

Mondloch, James Lorin. 1981. *Voice in Quiche-Maya*. PhD Dissertation, State University of New York at Albany.

Müller, Gereon. 2015. "Structure Removal: An Argument for Feature-Driven Merge." *Glossa: A Journal of General Linguistics* 2 (1): 1–35.

Murguia, Elixabete. 2004. *Syntactic Identity and Locality Restrictions on Verbal Ellipsis*. PhD Dissertation, University of Maryland, College Park.

Murphy, Andrew. 2016. "Subset relations in ellipsis licensing". *Glossa: A Journal of General Linguistics* 1(1): 44. 1-34.

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*. Berlin: Language Science Press. 313-336.

Nakao, Chizuru. 2009. *Island Repair and Non-Repair by PF Strategies*. PhD Dissertation, University of Maryland at College Park.

Napoli, Donna Jo. 1983. "Missing complement sentences in English: A base analysis of Null Complement Anaphora". *Linguistic Analysis* 12: 1-28.

Norris, Mark. 2014. *A theory of nominal concord*. PhD Dissertation, University of California at Santa Cruz.

Norris, Mark. 2017. "Description and analyses of nominal concord (Pts I-II)". *Language and Linguistics Compass* 11.

Nunberg, Geoffrey. 1993. "Indexicality and deixis". *Linguistics and Philosophy* 16: 1-43.

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*. John Benjamins. 215-236.

Nykiel, Joanna. 2019. "Polish". In Jeroen van Craenenbroeck & Tanja Temmerman (eds.), *The Oxford Handbook of Ellipsis*. Oxford: Oxford University Press.

Oku, Satoshi. 1998. *A theory of selection and reconstruction in the Minimalist Program*. PhD Dissertation, University of Connecticut at Storrs.

Ordóñez, Francisco. 1995. "The antipassive in Jacaltec: A last resort strategy." *CatWPL* 4: 329-343.

Ordóñez, Francisco. 1997. *Word order and clause structure in Spanish and other romance languages*. Doctoral Dissertation, CUNY.

Ortega-Santos, Ivan, Masaya Yoshida & Chizuru Nakao. 2014. "On ellipsis structures involving a wh-remnant and a non-wh-remnant simultaneously". *Lingua* 138: 55-85.

Otaki, Koichi, Koji Sugisaki, Noriaki Yusa, & Masatoshi Koizumi. 2013. "The parameter of argument ellipsis: the view from Kaqchikel". In Michael Kenstowicz (ed.), *Studies in Kaqchikel Grammar*. MIT Working Papers in Linguistics: 153-162.

Park, Dongwoo. 2017. *When does ellipsis occur, and what is elided?*. PhD Dissertation, University of Maryland at College Park.

Patal Majzul, Filiberto, Lolmay Pedro García Matzar, & Carmelina Ixchel Espantzay Serech. 2000. *Rujunamaxik Ri Kaqchikel Chi': Variación Dialectal En Kaqchikel*. Guatemala: Editorial Cholsamaj.

Patal Majzul, Filiberto. 2013. *Rusoltzij ri Kaqchikel: Diccionario Kaqchikel Estándar*. Guatemala City: Editorial Maya' Wuj.

Pearson, Matt. 2005. "The Malagasy Subject/Topic as an A′-Element." *Natural Language and Linguistic Theory* 23: 381–457.

Pearson, Matt. 2012. "Aspect and Voice Selection in Malagasy: Initial Observations". In T. Graf, D. Paperno, A. Szabolsci, & J. Tellings (eds.), *Theories of Everything: In Honor of Ed Keenan.* UCLA Working Papers in Linguistics 17, UCLA Department of Linguistics. 337-347.

Perlmutter, David M. 1971. *Deep and Surface Structure Constraints in Syntax*. The Transatlantic Series in Linguistics. New York: Holt, Rinehart, and Winston.

Pesetsky, David. 1987. "Wh-in-Situ: Movement and Unselective Binding." In Eric J. Reuland and Alice G. B. ter Meulen (eds.), *The Representation of (In)Definiteness.* Cambridge MA: MIT Press.

Pesetsky, David. 2017. "Exfoliation: Towards a Theory of Clause Size." Handout for talk presented at Workshop on Quirks of Subject Extraction.

Pesetsky, David. 2019. *Exfoliation: Towards a derivational theory of clause size*. Ms, MIT. Version 1.6; March 24th, 2019. Available at https://ling.auf.net/lingbuzz/004440.

Pesetsky, David. 2021. *Exfoliation: Towards a derivational theory of clause size*. Ms, MIT. Version 3; February 20th, 2021. Available at https://ling.auf.net/lingbuzz/004440.

Pesetsky, David & Esther Torrego. 2001. "T-To-C Movement: Causes and Consequences". In Michael Kenstowicz (ed.), *Ken Hale: A Life in Language*. Cambridge, MA: MIT Press. 355-426.

Pesetsky, David & Esther Torrego. 2007. "The syntax of valuation and the interpretability of feaures". In S. Karimi, V. Samiian, & W. K. Wilkins eds., *Phrasal and Clausal Architecture.* Syntactic Derivations and Interpretation. Amsterdam and Philadelphia: John Benjamins, 262-94.

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.

Picallo, M. Carme. 2009. "Gender and number in Romance". *Lingue e linguaggio* 7: 47-
66.

Polian, Gilles. 2017. "Morphology". In Judith Aissen, Nora England, & Roberto Zavala,
(eds.), *The Mayan Languages*. London: Routledge.

Polinsky, Maria. 2016a. *Deconstructing Ergativity: Two Types of Ergative Languages and*
Their Features. Oxford: Oxford University Press.

Polinsky, Maria. 2016b. "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. 2017. "Antipassive". In Jessica Coon, Diane Massam, & Lisa Travis
(eds.), *The Oxford Handbook of Ergativity*. 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.

Potsdam, Eric. 1996. "English verbal morphology and VP ellipsis". In *Proceeding of the*
27th *Meeting of the North Eastern Linguistic Society* 27. GLSA, University of
Massachusetts. Amherst. 353-368.

Potsdam, Eric. 2003. "Evidence for Semantic Identity under Ellipsis from Malagasy
Sluicing". In Makoto Kadowaki and Shigeto Kawahara (eds.), *The Proceedings of the*

33rd Meeting of the North East Linguistic Society. Amherst, Mass.: GLSA, University of Massachusetts at Amherst, 285-302.

Potsdam, Eric. 2007. "Malagasy sluicing and its consequences for the identity requirement on ellipsis". *Natural Language and Linguistic Theory* 25: 577-613.

Potsdam, Eric. 2018. "Exceptives and ellipsis". In Shay Hucklebridge & Max Nelson (eds.), *Proceedings of the 48th Meeting of the North East Linguistic Society*, volume 1. Amherst: GLSA. 259-268.

Potsdam, Eric. 2019. "Exceptives: An under-appreciated ellipsis construction". In Eszter Ronai, Laura Stigliano, & Yenan Sun (eds.), *Proceedings of the 54th Meeting of the Chicago Linguistics Society*. Chicago: CLS. 435-450.

Potsdam, Eric & Maria Polinsky. 2019. "Clausal phrases and exceptives". Paper presented at GLOW 42, Oslo, Norway.

Prado, Marcial. "El género en español y la teoría de la marcadez". *Hispania* 65: 258-266.

Preminger, Omer. 2009. "Breaking Agreements: distinguishing agreement and clitic-doubling by their failures". *Linguistic Inquiry* 40(4): 619-666.

Preminger, Omer. 2011. *Syntactic identity in ellipsis favors feature-sharing over feature checking.* Ms, MIT.

Preminger, Omer. 2014. *Agreement and Its Failures*. Cambridge, MA: MIT Press.

Preminger, Omer. 2020. "Functional structure in the noun phrase: revisiting Hebrew nominals". *Glossa: A Journal of General Linguistics*, 5(1), 68. DOI: http://doi.org/10.5334/gjgl.1244

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, & Jan Svartvik. 1972. *A Grammar of Contemporary English*. London: Seminar Press.

Rackowski, Andrea & Norvin Richards. 2005. "Phase Edge and Extraction: A Tagalog Case Study". *Linguistic Inquiry* 36(4): 565-599.

Ranero, Rodrigo. 2019a. "Voice Mismatches in Kaqchikel (Mayan) Sluicing". To appear in *Proceedings of WSCLA 24*. Available at https://ling.auf.net/lingbuzz/004865.

Ranero, Rodrigo. 2019b. *The eventive core is not special: A reply to Rudin (2019)*. Ms. University of Maryland, College Park.

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*. San Diego, CA: Academic Press. 37–62.

Rizzi, Luigi. 1997. "The Fine Structure of the Left Periphery". In Liliane Haegeman (ed.), *Elements of Grammar*. Springer: Dordrecht. 281-337.

Roberts, Ian. 1998. "Have/Be Raising, Move F, and Procrastinate". *Linguistic Inquiry* 29(1): 113-25.

Roca, Ignacio. 1989. "The organisation of grammatical gender". *Transactions of the Philolological Society* 87: 1–32

Rodrigues, Cilene, Andrew Nevins, and Luis Vicente. 2009. "Cleaving the Interactions between Sluicing and P-Stranding." In Torck Danièle and Wetzels W. Leo (eds.), *Romance Languages and Linguistic Theory 2006*. Amsterdam: John Benjamins. 245–70.

Rooryck, Johan & Erik Schoorlemmer. 2014. "Mistaken identity: ellipsis, mismatches and underspecification". Talk given at the *2014 Annual Meeting of the Linguistic Society of America*.

Rooth, Mats. 1992. "Ellipsis redundancy and reduction redundancy". In Steve Berman & Arild Hestvik (eds.), *Proceedings of the Stuttgart Ellipsis Workshop*. 1-26.

Rosen, Carol. 1976. "Guess What About." In A. Ford, J. Reighard, and R. Singh (eds.), *Papers from the 6th Meeting of the North Eastern Linguistic Society*. Montreal: Montreal Working Papers in Linguistics.

Rosenbaum, Peter Steven. 1965. *The Grammar of English Predicate Complement Constructions*. PhD Dissertation, MIT.

Rosenbaum, Peter Steven. 1967. *The Grammar of English Predicate Complement Constructions*. Cambridge, MA: MIT Press.

Ross, John Robert. 1967. "Constraints on Variables in Syntax." PhD dissertation, MIT. [Published in 1986 as Infinite Syntax! Norwood: Ablex].

Ross, John R. 1969a. "Guess Who?" In Robert Binnick, Alice Davison, Georgia Green, & Jerry Morgan (eds.), *Papers from the 5th Regional Meeting of the Chicago Linguistic Society*. Chicago: Chicago Linguistic Society.

Ross, John Robert. 1969b. "A Proposed Rule of Tree Pruning." In D. A. Reibel and S. A. Schane (eds.), *Modern Studies in English*. Prentice-Hall. 288–99.

Rudin, Deniz. 2019. "Head-Based Syntactic Identity in Sluicing". *Linguistic Inquiry* 50(2): 253-283.

Saab, Andrés. 2003. "Identidad Morfológica Estricta e Inserción Tardía". Talk given at *II Encuentro de Gramática Generativa*, Instituto en Lenguas Vivas "Juan Ramón Fernández".

Saab, Andrés. 2004. *El Dominio de la Elipsis Nominal en Español: Identidad Estricta e Inserción Tardía*. Master's Thesis, Universidad Nacional del Comahue.

Saab, Andrés. 2008. *Hacía una teoría de la identidad parcial en la ellipsis*. PhD Dissertation, Universidad de Buenos Aires.

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.

Saab, Andrés. 2015. "Ellipsis without truth". Ms., CONICET-UBA. Available at https://ling.auf.net/lingbuzz/002764.

Saab, Andrés. 2016. "Ineffable narratives in Spanish: Another case of overgeneration by e-GIVENness". *Probus* 28(2): 367-89.

Saab, Andrés. 2019. "Nominal ellipsis". In Jeroen van Craenenbroeck & Tanja Temmerman (eds.), *The Oxford Handbook of Ellipsis*. Oxford: Oxford University Press.

Saab, Andrés. 2021. *Ellipsis: Its Way from Syntax to Morphology*. Ms. CONICET-UBA. Version 6. Available at https://ling.auf.net/lingbuzz/004522

Sag, Ivan Andrew. 1976. *Deletion and Logical Form*. PhD Dissertation, MIT.

Sailor, Craig. 2014. *The Variables of VP Ellipsis*. PhD Dissertation, University of California at Los Angeles.

Saito, Mamoru & Keiko Murasugi.1990. "N'-deletion in Japanese". In Javier Ormazabal & Carol Tenny (eds.), *University of Connecticut working papers in linguistics 3*. Connecticut: University of Connecticut. 87-107.

Schoorlemmer, Erik. 2009. *Agreement, Dominance, and Doubling: The morphosyntax of DP*. Utrecht: LOT Publications.

Siddiqi, Daniel. 2009. *Syntax within the word: Economy, Allomorphy, and Argument Selection in Distributed Morphology*. Amsterdam: John Benjamins.

Silberman, Pamela Ann. 1995. *A Survey of the Use of Wi in Kaqchikel: Spoken and Written Language Norms*. MA Thesis, University of Texas at Austin.

Smith-Stark, Thom. 1978. "The mayan antipassive: Some facts and fictions". In Nora England (ed.), *Papers in Mayan linguistics*. The Curators of the University of Missouri. 169-187.

Soltan, Usama. 2016. "On the syntax of exceptive constructions in Egyptian Arabic". In Stuart Davis & Usama Soltan (eds.), *Perspectives on Arabic Linguistics XXVII*. Amsterdam: John Benjamins. 35-57.

Sprouse, Jon. 2005. "The Accent Projection Principle: Why the Hell Not?" In *U. Penn Working Papers in Linguistics, Volume 10.1*. Philadelphia.

Sprouse, Jon, Troy Messick, & Jonathan David Bobaljik. 2020. *Gender asymmetries in ellipsis: an experimental comparison of markedness and frequency accounts in English*. Ms. Available at: https://ling.auf.net/lingbuzz/005545

Stigliano, Laura. 2019. "P-stranding in ellipsis in Spanish does not arise from copular sources: evidence from non-exhaustive readings". To appear in *Proceedings of NELS 49*.

Stjepanovic, Sandra. 2012. "Two Cases of Violation Repair Under Sluicing." In Jason Merchant & Andrew Simpson (eds.), *Sluicing: Cross-Linguistic Perspective*. Oxford: Oxford University Press.

Stockwell, Richard & Deborah J. M. Wong. 2020. "Sprouting and the structure of except-phrases". In Mariam Asatryan, Yixiao Song & Ayana Whitmal (eds.), *Proceedings of the 50th Meeting of the North East Linguistic Society.*

Sudo, Yasutada & Giorgos Spathas. 2016. *Natural Gender and Interpretation in Greek.* Ms. University College London and Humboldt-Universität zu Berlin/Universität Stuttgart.

Sudo, Yasutada & Giorgos Spathas. 2020a. "Gender and interpretation in Greek: Comments on Merchant (2014)". *Glossa: a journal of general linguistics* 5(1): 129. 1-42.

Sudo, Yasutada & Giorgos Spathas. 2020b. "Gender on animal nouns in Greek". *Catalan Journal of Linguistics* 19: 25-48.

Tanaka, Hidekazu. 2011. "Syntactic identity and ellipsis". *The Linguistic Review* 28: 79-110.

Thoms, Gary. 2010. "Verb-floating and VP-ellipsis: Towards a movement account of ellipsis licensing". In Jeroen van Craenenbroeck (ed.), *Linguistic Variation Yearbook 2010.* Amsterdam: John Benjamins. 252-297.

Thoms, Gary. 2015. "Syntactic Identity, Parallelism and Accommodated Antecedents." *Lingua* 166: 172–98.

Thoms, Gary. 2019. "Varieties of English". In Jeroen van Craenenbroeck & Tanja Temmerman (eds.), *The Oxford Handbook of Ellipsis.* Oxford: Oxford University Press.

Toosarvandani, Maziar & Coppe van Urk. 2013. "The syntax of nominal concord: what ezafe in Zazaki shows us". In *Proceedings of NELS 43.* 209-220.

Vicente, Luis. 2019. "Sluicing and its subtypes". In Jeroen van Craenenbroeck & Tanja Temmerman (eds.), *The Oxford Handbook of Ellipsis*. Oxford: Oxford University Press.

Vostrikova, Ekaterina. 2019. "Compositional analysis for clausal exceptives". In *Proceedings of SALT* 29: 420-440.

Warner, Anthony. 1986. "Ellipsis conditions and the status of the English copula". *York Papers in Linguistics* 12. 153-172.

Wasow, Thomas. 1972. *Anaphoric relations in English*. PhD Dissertation, MIT.

Weir, Andrew. 2014. "Fragments and Clausal Ellipsis." PhD Dissertation, University of Massachusetts at Amherst.

Winkler, Susanne. 2016. "Ellipsis and Information Structure." In Caroline Féry and Shinichiro Ishihara (eds.), *The Oxford Handbook of Information Structure*. Oxford: Oxford University Press. 359–82.

Wood, Jim, Matthew Barros & Einar Freyr Sigurðsson. 2019. "Case mismatching in Icelandic clausal ellipsis". *Journal of Linguistics* 56 (2). 399-439. DOI: doi.org/10.1017/S0022226719000161.

van der Wurff, Wim. 2007. "Imperative Clauses in Generative Grammar: An Introduction". In Wim van der Wurff (ed.), *Imperative Clauses in Generative Grammar: Studies in Honour of Frits Beukema*. Amsterdam: John Benjamins. 1-94.

Zanuttini, Rafaella, Miok Pak, and Paul Portner. 2012. "A Syntactic Analysis of Interpretive Restrictions on Imperative, Promissive, and Exhortative Subjects." *Natural Language and Linguistic Theory* 30 (4): 1231–74.

Zagona, Karen. 1988. "Proper government and antecedentless VPs in English and Spanish". *Natural Language and Linguistic Theory* 6: 95-128.

Zocca, Cynthia. 2003. *O que não está lá. Um estudo sobre morfologia flexional em elipses*. Master's Thesis, Universidade Estadual de Campinas.

Zwicky, Arnold. 1981. "Stranded *to* and Phonological Phrasing". *Linguistics* 20: 3-57.