

VARASCHIN, Giuseppe. Unified accounts of English reflexives: history and prospects. *ReVEL*, edição especial, v. 17, n. 16, 2019. [www.revel.inf.br]

UNIFIED ACCOUNTS OF ENGLISH REFLEXIVES: HISTORY AND PROSPECTS

Abordagens unificadas de pronomes reflexivos no inglês: história e perspectivas

Giuseppe Varaschin¹

giuseppe.varaschin@gmail.com

ABSTRACT: Most of the early approaches to reflexivization in Mainstream Generative Grammar (MGG) assumed that a *single* constraint applied uniformly to all instances of reflexive pronouns in English. These theories provided, therefore, unified constraints, which did not distinguish, *inter alia*, different types of grammatical licensing conditions for reflexives according to their syntactic contexts. The main goal of this paper is to survey the landmarks in the history of these unified accounts. I focus particularly on the earliest and on the most prominent of these accounts: Lees and Klima (1963) and the Classical Binding Theory (respectively). These two theories exemplify two different strategies in facing counterexamples to their postulates: Enrich Structure or Enrich Principle. I will show that both of these strategies are insufficient and that a third one is needed: we must lexicalize the constraint on reflexives and make it *violable*. Assuming that, I briefly sketch my alternative unified account within the constructional framework of Culicover and Jackendoff's (2005) Simpler Syntax.

KEYWORDS: Binding Theory; reflexives; English Syntax; Simpler Syntax.

RESUMO: A maior parte das primeiras propostas sobre reflexivização na Gramática Gerativa Mainstream (GGM) assumiam que uma *única* restrição se aplicava uniformemente a todas as instâncias de pronomes reflexivos no Inglês. Essas teorias proporcionavam, desse modo, restrições unificadas, que não distinguiam, *inter alia*, diferentes tipos de condições de licenciamento para os reflexivos de acordo com seus contextos sintáticos. O principal objetivo desse artigo é revisar os principais marcos na história dessas abordagens unificadas. Focarei em especial no representante mais antigo e no mais proeminente dessas abordagens: Lees and Klima (1963) e a Teoria da Ligação Clássica (respectivamente). Essas duas teorias exemplificam duas estratégias distintas no trato de contraexemplos aos seus postulados: Enriquecer Estrutura ou Enriquecer Princípio. Mostrarei que essas duas estratégias são insuficientes e que uma terceira estratégia é necessária: devemos lexicalizar a restrição sobre os reflexivos e torná-la *violável*. Assumindo isso, o artigo expõe também o rascunho breve de uma abordagem unificada alternativa dentro do quadro da Sintaxe mais Simples de Culicover e Jackendoff (2005).

PALAVRAS-CHAVE: Teoria da Ligação; reflexivos; sintaxe do inglês; sintaxe mais simples.

¹ PhD student in the Federal University of Santa Catarina (UFSC).

INTRODUCTION

With the plausible exception of *wh*-words, English reflexive pronouns are solid candidates for the most hotly debated lexical items among the vast inventory of forms found in the world's languages. This situation is somewhat surprising, since, before a recent profusion of debate in the latter half of the twentieth century, traditional grammars were remarkably shy about the topic. As Reuland (2014: 1) notes, for most of the history of inquiry on language, reflexivity was presented as "a relatively marginal phenomenon". All Jespersen had to say about it in his nearly 400 page grammar of English can be summarized in the following:

When the subject and object are identical, we use for the latter a so-called reflexive pronoun, formed by means of *self* e.g. *I defend myself*. [...] The reflexive pronouns are also used after prepositions: He looked at himself in the glass. (Jespersen 1933: 111-112)

Given the amount of ink spent over the last fifty years, the above quote can sound almost naive. But it is not obvious why a particular subset of pronouns in a language should be a major concern for linguists. Far from suggesting a linear progress towards a consensus, the sheer number of diverse proposals in recent decades is a testament to how much there is to be discovered even in apparently banal territories of grammar.

Mysteries turn up in a wealth of puzzling facts. Along with well-behaved contrasts like the one between (1) and (2), which clearly conform to Jespersen's intuition (and to virtually all accounts since then), one can readily find data in which reflexives do not appear to signal covaluation between bearers of grammatical functions within a clause (e.g. subjects and objects), such as (3).²

- (1) *Janet* loves *herself*. (cf. **Janet* loves *her*.)
- (2) a. *An old boyfriend of *Janet* still loves *herself*.
 b. **Janet* told me that Brad loves *herself*.
 c. **Janet's* mom loves *themselves*.

² In order to avoid the intrusion of particular theoretical assumptions in the presentation of the data, I adopt the notation in Safir (2004), signaling semantic covaluation between expressions by means of italics (unless a particular theory involving indexing is discussed). As usual when it comes to studies of anaphora, judgments refer to string-interpretation pairs (and not merely to strings themselves).

- (3) a. “[...] *she* is not *herself* anymore.”
 b. “[...] a little part of *me* still hates *myself* for letting that happen.”
 c. “*Pears* [...] believes some customers would be similar to *herself*.”
 d. “[...] it was *herself* that *she* hated, and not the Little Sister.”
 e. “[...] the picture of *himself* that *he* gave his girlfriend will remind her with what she originally fell in love [...]”
 f. “[...] *she* found herself experimenting with materials and soon had a vision to create a clothing line that would not only give her an emotional outlet, but would empower other young women like *herself*.”
 g. “This post is an almost accurate description of myself.”³

It is mostly around problematic cases like (3) that research on English reflexives has gravitated over the last fifty years. In (3a), identity between subject and object is explicitly asserted not to hold. In (3b), like in the ill-formed (2a), the antecedent is somehow embedded within the subject. In (3c)-(3f), reflexives and their antecedents are not even clausemates, since there is a finite clause boundary separating them, as also occurs in (2b). Example (3g) is akin to (2c) in that there is no antecedent in the sentence at all. Besides that, in most of these cases, the complementarity between reflexives and pronouns exemplified by the contrast in (1) breaks down: replacing the reflexive for a pronoun would keep meanings constant.

Most of the early approaches in Mainstream Generative Grammar (MGG) – with the notable exception of Postal (1971) – followed Jespersen in assuming that constraints on reflexives applied uniformly to all instances of reflexives in a language. These theories provided, therefore, *unified* constraints, which did not distinguish, *inter alia*, different types of grammatical licensing conditions for reflexives according to their syntactic contexts.

However, since then, the idea that a unified interpretive constraint applies to all reflexives has become a minority position for a good reason (cf. Pollard and Sag 1992; Reinhart and Reuland 1993): even if we restrict ourselves to an overstudied language like English, the sheer diversity of reflexives exemplified in (3) makes unified accounts too complicated. The fact that reflexives appear in all sorts of diverse

³ The data in (3) is drawn from the iWeb Corpus (<https://www.english-corpora.org/iweb/>).

configurations makes the idea of specifying *disjunctive* licensing conditions very tempting.

Accordingly, syntacticians sought to state rules which explicitly guaranteed a schism between well-behaved law-abiding reflexives (such as (1) and (2)) and fancy-free anything-goes reflexives (like the ones in (3)). The former have to meet grammatically identifiable requirements while the latter are explicitly liberated from any such obligations (i.e. they are *exempt*). The idea that grammar itself specifies whether reflexives are subject to grammatical requirements or not is what I call the Two Reflexives Hypothesis (TRH). The popularity of the TRH since the late 1980s reveals that most researches have, in effect, abandoned the formulation of unified constraints on anaphoric forms.

The main goal of this paper is to survey the landmarks in the history of unified accounts for English reflexives in Mainstream Generative Grammar (MGG). I will also suggest a somewhat heterodox strategy that enables a formulation of a unified account which avoids the problems posed by examples like (3). This heterodox strategy involves *lexicalizing* the constraint on reflexives and admitting that this constraint can be *violated*. This work is a summary of Varaschin (2019), where the argument and background for this analysis is spelled out in greater detail.

Section 1 covers the treatment of reflexives in Lees and Klima (1963) — i.e. the *earliest* account of English reflexives in MGG. Section 2 presents the Classical Binding Theory of Chomsky (1981, 1986) and others — i.e. the most *prominent* treatment of reflexives in MGG. Both of these accounts exemplify two different strategies for proposing unified accounts of English reflexives. In Section 3, I briefly sketch (in a kind of programmatic style) my alternative unified account within the constructional framework of Culicover and Jackendoff's (2005) *Simpler Syntax*. I lay out some concluding remarks.

1 LEES AND KLIMA (1963)

Lees and Klima (1963) launched the first systematic investigation of anaphora in the tradition of Mainstream Generative Grammar (MGG).⁴ Their work is a useful

⁴ Here and throughout this paper I will use the term *anaphora* (and its cognates) as a general label for the phenomena of intended semantic covaluation between linguistic expressions. *Semantic covaluation* covers two different kinds of interpretive dependencies: coreference and bound-variable

starting point for my discussion because it presents, in its virtues and flaws, a particularly clear statement of a unified approach to reflexives. Lees and Klima's (1963) motivation was to employ the (then recently developed) machinery of transformational analysis to shed light into regularities that were not captured by semantically-based theories of anaphora. Their main hypothesis was that the distribution and interpretation of anaphoric forms (reflexives and pronouns) were subject to *syntactic rules* whose formulation required some of the tools put forth by Chomsky (1957).

Initially, these rules were taken to be transformations – i.e., technically, rules that map strings with a given constituent structure (phrase-markers) into new strings with derived constituent structure (Chomsky 1957: 44). Two transformations were proposed to account for the interpretation and distribution of anaphoric forms in English: (i) *reflexivization*, which converted the second of two identical nominals (Nom) within a simplex sentence into a reflexive SELF form⁵, and (ii) *pronominalization*, which converted the second of two identical nominals in different sentences into a bare pronoun. Since Lees and Klima (1963) proposed a *single* rule to account for *each* anaphor in English (one for reflexives and one for pronouns) their approach counts as a unified one. The rules were originally stated as follows:

(4) **Reflexivization Rule (RR)** (Lees and Klima, 1963: 23):

$X - \text{Nom} - Y - \text{Nom}' - Z \rightarrow X - \text{Nom} - Y - \text{Nom}' + \text{Self} - Z$

Where $\text{Nom} = \text{Nom}'$ and Nom and Nom' are within the same simplex sentence.

relations (cf. Postal 1971 and Reinhart, 1983). Both are customarily signaled, since Chomsky (1965), by means of coindexing – a practice I mostly abandon in favor of italics. An *anaphor* (or an *anaphoric form*) is simply a linguistic form lexicalized for the expression of anaphora. My use of the term is, thus, closer to the traditional grammarian's use than to the terminological practice popularized by Chomsky (1981), who used “anaphor” to name a class of NPs which includes only reflexives (e.g. *himself*) and reciprocals (e.g. *each other*). We can think of Chomsky's (1981) “anaphors” as NPs which are *necessarily anaphoric* (in my sense): they *must* be covalued to another expression in order to be assigned a meaning. A property which identifies these “necessarily anaphoric” NPs is their *deictic deficiency* – i.e. they can't be used as deictics (cf. Safir 2003).

⁵ I am assuming the tripartite typology of anaphoric forms proposed in Reinhart and Reuland (1993), which is defined in terms of two features: R (referential independence) and REFL (reflexivizing function). Regular pronouns are [+R, -REFL]. Reflexive pronouns in English are [-R, +REFL]. English has no form whose feature profile is [-R, -REFL], but such forms do exist in other languages, usually corresponding to SE (i.e. simplex expression) reflexives. These are usually morphologically simpler than [+REFL] reflexives. An example would be the Dutch anaphor *zich* (Everaert 1986).

(5) **Pronominalization Rule (PR)** (Lees and Klima, 1963: 23):
$$X - \text{Nom} - Y - \text{Nom}' - Z \rightarrow X - \text{Nom} - Y - \text{Nom}' + \text{Pron} - Z$$

Where $\text{Nom} = \text{Nom}'$ and Nom is in a matrix sentence while Nom' is in a constituent sentence embedded within that matrix sentence.

Non-pronominal NPs were taken to be present in the underlying kernels for sentences with anaphoric forms. The transformational outlook preserved, thus, the view expressed in traditional grammar that pronominal expressions are words that “replace” nouns, as the derivation in (6) illustrates:⁶

(6) John loves John

$$\text{RR} \rightarrow \text{John loves himself}$$

This position became unpopular after Jackendoff (1969) and Chomsky (1972) proposed a full-fledged interpretive system for coreference. They argued that pronouns and reflexives should no longer be *derived* from underlying fully specified NPs, but base-generated as real lexical items. Rules like (4)-(5) were supplanted by rules of *construal*: i.e. rules which regulate the *interpretation* of NPs, as schematized in (7) (adapted from Jackendoff 1969: 45):

(7) $\left[\begin{array}{c} \text{NP}^2 \\ \text{+/- pron/refl} \end{array} \right]$ is interpreted as +/- covalued to NP^1 in the context X.

⁶The examples in (i) show the interaction between the RR and other transformations, such as the Identity Erasure Transformation (IET) (Rosenbaum 1967), later to be known as Equi-NP Deletion, and Conjunction Reduction (CR) (Chomsky 1957):

- (i) a. John tried to John love John
 $\text{RR} \rightarrow \text{John tried to John love himself}$
 $\text{IET} \rightarrow \text{John tried to love himself.}$
- b. John prefers John to Bill
 $\{\text{RR} \rightarrow \text{John prefers himself to Bill; John prefers Mary to Bill}\}$
 $\text{CR} \rightarrow \text{John prefers himself and Mary to Bill}$

The transformations have to be applied in precisely the order given above in order to yield the right results. The CR transformation in (i-b) is an instance of a *generalized transformation*: a transformation that operates on more than one string to yield a derived structure.

The interpretive stance has been close to a consensus since these early formulations – with Lidz and Idsardi (1998) and Hornstein (2001) as virtually the only outliers. I will also assume a version of it here. Hence, my use of terms such as “reflexivization”, unless otherwise indicated, refers to rules of the kind in (7).

There were some important empirical reasons for this shift towards interpretive theories of anaphora. One, which is implicit in Jackendoff’s (1969) formulation of (7), was that rules of construal can specify *disjoint reference* just as easily as coreference. This seemed necessary to account for the meaning of expressions like *some-x else* (Culicover and Jackendoff 1995) and for the effects of what later came to be called Conditions B and C of the Binding Theory (Chomsky 1981; Reinhart 1983). Moreover, assuming interpretive rules avoids positing infinitely recursive underlying structures like (8b) for (8a) (Jackendoff, 1969: 42) and synonymy between (9a) and (9b) (Jackendoff 1968: 433).

- (8) a. The man who deserves it will get the prize he wants.
 b. The man who deserves [the prize [which the man [who ...]] wants] will get the prize [the man [who deserves the prize [which ...]]] wants.
- (9) a. *Everyone* likes *himself*.
 b. Everyone likes everyone.

Both of these difficulties arise in a system like Lees and Klima’s (1963), where coreference between two surface forms is defined derivationally as identity on an underlying level of structure.

However, not all of the obstacles Lees and Klima (1963) encounter are directly related to their outmoded transformational mindset. Some of them stem from the very *simplicity* of their rule system. Simple rules like (4) and (5) require somewhat artificial assumptions to handle data like (10)-(12):

- (10) a. *The men* found a smokescreen around *them*.
 b. *The men* threw a smokescreen around *themselves*.
- (11) a. *John* saw many pictures of *him*.
 b. *John* saw many pictures of *himself*.

- (12) a. **Mary's father supported herself.*
 b. *Mary's father supported her.*

Assuming that all of the sentences in (10)-(11) are simplex, only reflexives are expected to occur. The fact that both reflexives and pronouns are equally acceptable is, thus, problematic. In line with most of the literature up to Huang (1983), Lees and Klima (1963) take complementary distribution between anaphoric forms to be the norm: where a reflexive can occur, a pronoun cannot occur (and *vice versa*). The assumption of complementarity derives from the fact that the environments for reflexivization and pronominalization are stated so as to be *mutually exclusive*.

The contrast in (12) presents an even more serious difficulty, because it is exactly the *opposite* of what RR and PR apparently predict. Note that the structural conditions for both of these rules – as was the case for transformations in general in early MGG (Chomsky 1957) – do not refer to the *constituent structure* of the strings to which they apply. All that is required for the RR in (4), for example, is that two nominals cohabit the same simplex sentence domain – a condition which (at first sight) is met in (12a). No other structural relations (e.g. c-command) were imposed at this stage, because the notion of domain relevant for anaphora was, basically, the simplest one imaginable: the (simplex) sentence.

In order to circumvent these problems while keeping their rules intact, Lees and Klima (1963) had to attribute richer (and largely unmotivated) covert structures to the strings (10)-(12). To account for the non-complementarity within locative PPs illustrated in (10), Lees and Klima (1963) proposed that (10a) and (10b) result from two distinct underlying derivations, exemplified below:

- (13) a. {The men found a smokescreen; A smokescreen was around the men}
 → The men found a smokescreen which is around the men
 PR → The men found a smokescreen which is around them
 → The men found a smokescreen around them
- b. The men threw a smokescreen around the men.
 RR → The men threw a smokescreen around themselves

According to Lees and Klima, the surface similarity between (10a) and (10b) turns out to be misleading, because only in (10b), i.e. in (13b), the nominal which is

target for reflexivization is actually part of the same simplex sentence as its antecedent.

For cases like (11a), Lees and Klima (1963, pg. 23) suggest the derivation (14a), which is similar to (13a). Each predication is, as was usual at the time, encoded as a separate kernel. However, this only accounts for pronouns and fails to explain how reflexives appear in (11b). As Lees and Klima (1963: 24) recognize, this is a particularly tricky issue because the pair in (11), unlike the one in (10), consists of strings which are identical, minus the anaphoric forms.⁷ The only story compatible with their framework would be to claim a kind of structural ambiguity. This requires enriching the base component to yield, along with the kernels in (14a), more complex kernels like (14b):

- (14) a. {John saw many pictures; The pictures are of John}
 → John saw many pictures which are of John
 PR→ John saw many pictures which are of him
 → John saw many pictures of him
- b. John saw many pictures of John
 RR→ John saw many pictures of himself

For the genitives in (12), the proliferation of underlying structures is even more telling. In order to explain how reflexivization is unacceptable (and pronominalization is acceptable) in an apparently “local” context like (12), Lees and Klima (1963) propose to treat genitives in English as transformational in origin, deriving from full sentences: e.g. *Mary’s father* is derived from *the father that Mary has* (which is itself a derived phrase-marker). In the latter, *Mary* is embedded within a relative clause. This entails that there is a point in the derivation of (12) in which *Mary* is not within the same sentence as the direct object of the verb *support*. If this point is the point where RR and PR apply, then only the PR could go through, because the two nominals which refer to Mary would not be in the same sentence:⁸

⁷ This is one of the motivations for Postal’s (1971, pg. 188) *late reflexivization* rule, which inaugurates the Two Reflexives Hypothesis. Unlike the standard reflexivization rule in (4), late reflexivization applies *after pronominalization* and whatever rule reduces the relative clause to a simple PP in (13a) and (14a).

⁸ Needless to say, even though the proposed solutions for (10)-(12) technically work, they are extremely artificial and theoretically dubious. They essentially rely on transformational mechanisms

- (15) {The father supported Mary; Mary has a father}
 → The father that Mary has supported Mary
 PR→ The father that Mary has supported her
 →Mary's father supported her

Examples like (10)-(12) are still nowhere near as complex as examples like (3). Nonetheless, Lees and Klima (1963) had to resort to a great number of nontrivial (and somewhat arbitrary) assumptions in order to explain them. Just to give a taste of how inadequate the theory is upon confrontation with a more realistic set of data, consider (16) and Heim's (1998) example in (17):

- (16) a. *Bill* knows that this picture of *herself*, Mary will like.
 b. *Bill* knows that this picture of *himself*, Mary will like.
 (17) *Everyone* thinks that only *he* can hear *him* sing in the bathroom.

In order to explain how Lees and Klima's system could yield a string like (16a), one would have to stipulate a kernel *Mary likes this picture of Mary*, to which RR would apply, as in (14b). A topicalization transformation (which "moves" the picture-NP to the front of the embedded clause) could only apply *after* RR. However, precisely the *opposite* ordering would have to be stipulated to account for (16b), where the picture-NP reflexive takes the *matrix* subject as its antecedent. If RR applied before topicalization, the antecedent of the reflexive could only ever be the embedded subject *Mary*, as in (16a).⁹

Heim's example (17), besides recapitulating the problem of quantified antecedents illustrated in (9), is a counterexample to Lees and Klima's theory, because *he* and *him* are covalued and belong to the same clause. This violates the PR. The reason why the sentence is acceptable, according to Heim, is because the pronouns, though covalued, are not actually dependent on *each other* — they are both

which were shown to be excessively powerful (Jackendoff 1969; Emonds 1970; Peters and Ritchie, 1973) – e.g. the rule in (14) which transforms a complex NP with a relative clause into an NP with a genitive subject. Furthermore, even if these language-specific and non-structure preserving transformations are accepted, the proposed ordering of these rules with respect to the RR and PR is still entirely *ad hoc*.

⁹ Moreover, it is unclear to which string RR could apply in (16b), since it is implausible that *Bill knows that this picture of Bill* is construed as a simplex sentence anywhere in the derivation.

codependent on the quantifier subject. Lees and Klima's transformational system (as well as simple interpretive theories) has no way to capture these dependencies.

Anachronisms aside, Lees and Klima's (1963) proposals are useful for my exposition because they flaunt a dilemma which affects all unified approaches to reflexives since then. In light of potential counterexamples to simple unified principles such as the RR, theorists can take one of the two following strategies:¹⁰

- (i) **Enrich Structure:** claim that the apparent counterexamples are not *really* counterexamples, but cases where there is a mismatch between overt and covert structure (where the latter in fact *confirms* the favored principle); or
- (ii) **Enrich Principle:** revise the principle and make it more complex (which either burdens the learning procedure, if the principle is learned, or UG, if the principle is innate).

Lees and Klima (1963) produced the perfect example of a simple unified theory because they opted for (i) in every case. This allowed them to keep their simple rule system intact in the face of diverse data – but only at the cost of having to postulate complex derivational mechanisms. Lees and Klima never considered the introduction of an additional structural relation (e.g. c-command) in order to refine their notion of domain to handle (7)-(9). Rather, they stuck to their crude definition of domains as simplex sentences and proceeded to proliferate complex covert structures for cases where their simple notion of domain apparently fails. This overall strategy, which is a corollary of syntactocentrism (cf. Culicover and Jackendoff 2005), is common to diverse treatments of anaphora within MGG. We will now turn to an approach which opted mainly for the strategy stated in (ii) above: the Classical Binding Theory (CBT).

2 THE CLASSICAL BINDING THEORY (CBT)

Even though the appeal to abstract syntactic structure to avoid counterexamples remained fashionable in MGG, it was generally agreed that the complex derivational mechanisms required by Lees and Klima (1963) were excessive

¹⁰ A third alternative would be to restrict the scope of the original principle and to propose a second principle to account for the remaining cases. This entails trading a unified approach for the Two Reflexives Hypothesis inaugurated by Postal (1971).

and unmotivated. In accordance with (ii), a more nuanced view of the principles responsible for reflexives was called for. Instances of this tactic were Jackendoff's (1969) Thematic Hierarchy Condition, Langacker's (1969) primacy relations (*precede* and *command*), Chomsky's (1973) Conditions framework and Reinhart's (1976) c-command requirement. All of these were attempts to impose *further* constraints on anaphora beyond what is suggested by the reflexivization and pronominalization rules (4)-(5), while still maintaining a unified approach.

This line of development crystalized in works such as Chomsky (1981, 1986), Huang (1983) and Reinhart (1983). These proposals have enough in common to merit a single label: they are all versions of what I call the Classical Binding Theory (CBT). Before getting into details, it is important to keep in mind that the CBT is conceptually a very different theory from Lees and Klima's. First, it is fully interpretive – anaphoric forms are taken to be real lexical items, and not spell-outs of transformations. Second, the CBT fits into a theory which allows recursive phrase-structure rules, as in Chomsky (1965). Third, semantic covaluation is signaled by means of syntactic coindexing, and not by identity of phrase-markers in covert structures – thereby avoiding the problems raised by (8) and (9). Fourth, the CBT, as its name suggests, is primarily concerned with *binding* and only derivatively concerned with the *distribution* of anaphoric forms.

The binding relation is defined below (Chomsky 1981: 184):¹¹

- (18) α is **bound** by β iff α and β are coindexed and β c-commands α ;
- a. β **c-commands** α iff neither α nor β dominate the other and the first branching node that dominates β also dominates α .

¹¹ Insofar as binding is syntactically defined, distributional or collocational effects are expected. (18) also suggests binding is semantically distinct from mere coreference – which is either not defined in narrow syntax (Reinhart 1983) or indicated by coindexing without c-command (Fiengo and May 1994). This difference is exemplified in (i), where sentences are paired with their logical representations in the usual λ -notation:

- (i) a. *Every man* thinks *he* is clever. (binding)
 Every man ($\lambda x. x$ thinks that x is clever)
- b. The woman who loves *John* thinks that *he* is clever. (coreference)
 [The woman ($\lambda y. y$ loves John)] ($\lambda x. x$ thinks z is clever) ($z = \text{John}$)

There's no binding in (ib) because *John* does not c-command the pronoun. In that case, the pronoun acts semantically as a free variable whose value must be contextually supplied. Note, however, that the concept of binding in (18) is still purely syntactic, and leaves open many details of how it should be semantically interpreted. See Reinhart (2006: chap. 4) for an illuminating discussion of these issues.

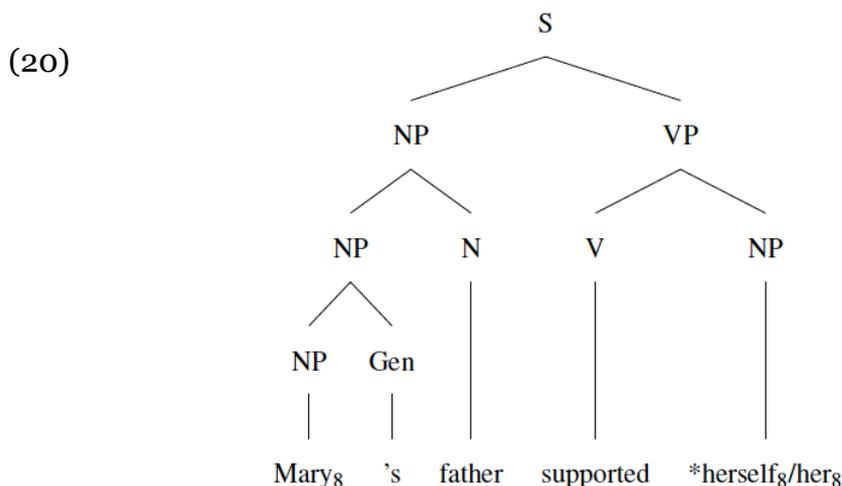
The CBT is essentially a theory about how each type of NP – reflexives, reciprocals, pronominals, etc. – must be, can be or shouldn't be bound. The principles of the CBT relevant for reflexives and pronouns – Binding Conditions A and B – have subtly different incarnations, but they all agree on the following (Chomsky 1981, 1986; Huang 1983; Reinhart 1983):

(19) **Classical Binding Theory** (CBT) (Chomsky 1986: 166)

a. **Condition A:** A **reflexive** is bound in a local domain.

b. **Condition B:** A **pronoun** is free (i.e. not bound) in a local domain.

Aside from a domain condition – which parallels to Lees and Klima's (1963) "simplex sentence" proviso – (19a) introduces an additional requirement for reflexives: they must be bound, and, therefore, c-commanded by their antecedents.¹² Even though this adds further complexity to the rule system, it allows a reduction of structure in the explanation of the pattern with genitives in (12), whose syntactic representation is sketched out below:



There is no need to resort to a complex derivation here. A reflexive is ruled out in (20) because its putative binder *Mary* does not c-command it. A pronoun is fine because pronouns are only subject to a negative requirement: they must not be locally bound. Since the pronoun in (15) is not bound at all, the structure is fine.

¹² The concept of domain invoked in (19) varied among versions of the CBT. I will address this issue below. The notion of c-command is taken to be independently necessary to account for other phenomena of language, such as long-distance dependencies, government and scope.

reduce derivational and structural complexity by incorporating more conditions into its rules.

Though the CBT was empirically and conceptually superior to previous theories (like Lees and Klima's), it still struggled to provide a satisfactory account for many phenomena. The non-complementarity of reflexives within subjectless NPs illustrated in (11), repeated below as (22), is a case in point:

- (22) a. *John* saw many pictures of *him*.
 b. *John* saw many pictures of *himself*.

The crucial problem here lies on the notion of “local domain”, which was deliberately left unspecified by the formulation in (14). Like Lees and Klima, most variants of the CBT also take complementarity between reflexives and pronouns to be the norm: the domain in which reflexives must be bound and the domain in which pronominals cannot be bound should be one and the same. If this domain is, for instance, a governing category as defined in Chomsky (1981: 211), (22a) lacks an explanation:

- (23) **Binding domains** are governing categories, where
 a. γ is a **governing category** for α if and only if γ is the minimal category containing α , a governor of α , and a SUBJECT accessible to α .¹⁴

Given (23), reflexives should be bound and pronouns should be free in the domain of the nearest (accessible) SUBJECT. This is apparently falsified by (22a). The dilemma (22) raises is this: how can a *single* notion of domain for reflexives and pronouns be made compatible with contexts in which they occur in free variation, such as (22)?

b. *The men* threw a smokescreen around *them*.

Moreover if external arguments move to [Spec, IP] (Koopman and Sportiche 1991), (21) demands that Binding Conditions A and B be checked *before* such movement takes place. This runs counter to the commonly held view that the Binding Theory applies at LF.

¹⁴ The concept of SUBJECT is introduced in Chomsky (1981: 209) as a technical notion intended to cover both subjects in the usual sense and AGR (the bundle of agreement features in verbal inflections). The notion of accessibility is explored below.

Chomsky's (1986: 67) answer is an instance of the strategy which proliferates covert structure in order to keep the rule system intact. Assuming (23), he argues that the NPs in (22) contain null pronominal (PRO) subjects. When these local PRO subjects are not coindexed to the anaphoric NPs, a pronoun is required. When they are coindexed, a reflexive is mandatory:

- (24) a. John₈ saw many [_{NP} PRO₂ pictures of him₈]
 b. John₈ saw many [_{NP} PRO₈ pictures of himself₈]

There are fair reasons to be suspicious about this analysis. First, it presupposes that NPs have PRO subjects – a position which was shown, on quite independent grounds, to be untenable (cf. Williams 1985). Second, even if PRO subjects are tolerated, it is unclear how they are supposed to be interpreted in structures like (24). As Lebeaux (1985: 347) notes, there is no detectable meaning difference between pairs like (24a)-(24b). There is no suggestion, for instance, that the pictures in (24a) (where PRO is not coindexed with *him*) belong to some unspecified person, whereas in (24b) they belong to John. PRO and its index appear to be semantically inert – they only function as *ad hoc* devices to save the CBT.

The only alternative compatible with the spirit of the CBT is to devise two distinct notions of domain: one for reflexives and one for pronouns. The domain in which a pronoun must be free has to be, in some sense, “smaller” than the domain in which a reflexive must be bound. This was more or less the gist of Huang's (1983) modification of Chomsky's (1981) system. Huang states that only governing categories for *reflexives* have to contain an accessible SUBJECT in the sense of (25a). He also redefines the concept of SUBJECT so as to include nominal heads (Huang 1983: 557-558):

- (25) γ is a **governing category** for α if and only if γ is the minimal category containing α , a governor of α , and a SUBJECT that, if α is necessarily anaphoric (e.g. a reflexive), is accessible to α .
 a. σ is **accessible** to α if and only if σ is in the c-command domain of α and assignment to α of the index of σ would not instantiate the *i*-within-*i* schema: i.e. the schema [_{ψ} ... δ_i ...]_{*i*} (Chomsky 1981: 212).

- b. The **SUBJECT** of a maximal phrase ψ is the subject of ψ (in the usual sense) or the nominal head of ψ .

According to (25), the minimal governing category for the pronoun in (22a) is the post-verbal NP *pictures of him*, since it contains the pronoun, its governor (i.e. the preposition *of*) and a SUBJECT – namely, the nominal head *pictures*, as per (25b). Since the pronoun is *not bound* in this domain, (22a) is fine. The domain for the reflexive in (22b), on the other hand, is predicted to be the whole S, because only S contains the reflexive, its governor and an *accessible* SUBJECT.¹⁵ Given that the reflexive in (22b) is bound in S (its local domain), (22b) also supports Huang's (1983) definitions.

Cases of non-complementarity across finite clause boundaries, like (26), fall under the same explanation (Huang 1983: 554):

- (26) a. *They* expected that books about *them* would be on sale.
 b. *They* expected that books about *themselves* would be on sale.

The governing category for pronoun in (26a) is the NP, for the same reason as in (22a). On the other hand, the governing category for reflexive in (26b) is the matrix sentence, since only the matrix sentence contains an accessible SUBJECT.

Huang's partitioning of binding domains thus accommodates into the CBT the non-complementarity of pronouns and reflexives within NPs without having to resort to unmotivated covert structures. To be sure, this is done at the expense of introducing further intricacies into the system, some of which have a dubious flavor (e.g. the stipulation that nominal heads are SUBJECTS in (25b)). Nonetheless, Huang's innovations were not unusual at the time and his account works reasonably well for the cases mentioned so far.

However, all of this is largely beside the point because there are real – and more severe – difficulties which jeopardize not only Huang's specific suggestions, but

¹⁵ The nominal head, though a SUBJECT in the technical sense defined in (25b), must not count as an *accessible* SUBJECT. For this to work, Huang (1983: 558) has to stipulate that the index of a nominal head percolates from its maximal projection. In that case, the N head of NP is never accessible to an anaphor α contained within that NP, since coindexing α with N in this configuration would instantiate the *i*-within-*i* schema in (25a): $[_{NP} N_i \alpha_i]_i$.

the CBT as a whole. The following assumptions are essential to all versions of the CBT (Pollard and Sag 1992: 263):

- (27) a. Reflexives must be c-commanded by their antecedents.
 b. Reflexives are never covalued to an NP beyond the domain of the nearest specified subject (cf. Chomsky 1973).
 c. Reflexives are never “discourse-bound”.

None of these propositions stand in light of a wider survey of data on English reflexives. Consider the following:¹⁶

- (28) a. That description of *himself* annoyed *DeGaulle* more than you know.
 b. Unflattering descriptions of *himself* have been banned by *LBJ*.
 c. The picture of *himself* in Newsweek shattered the peace of mind that *John* had spent the last six months trying to restore.
 d. *Albert* was never hostile to laymen who couldn’t understand what physicists like *himself* were trying to prove.
 e. *John* believes that Mary would never consider marrying a man less wealthy than *himself*.
 f. *Brad* warned Janet that she shouldn’t trust anyone but *himself*.
 g. *The men* said that the new recruits would be very much like *themselves*.
 h. *Max* boasted that the queen invited Lucie and *himself* for a drink.
 i. As for myself, I won’t be invited.
 j. Physicists like myself were never too happy with the parity principle.

The c-command requirement (27a) is violated in (28a)-(28c) and in (28h)-(28j). Cases (28c)-(28h) show, *contra* (27b), that reflexives *can* be covalued to NPs across specified subjects. Furthermore, this “long-distance” reflexivization is not

¹⁶ Most of the phenomena that contradict (27) were known at least since Jackendoff (1969) and Cantrall (1969). This shows that, more commonly than most would like to admit, proponents of the CBT had to sweep known data under the rug to make a unified theory seem plausible – a point made by Zribi-Hertz (1989: 703) and Reuland (2011: 44) alike. Example (28a) comes from Postal (1971: 188), (28b) comes from Jackendoff (1969: 40), (28c) comes from Pollard and Sag (1992: 278), (23-d) and (28h)-(28i) come from Ross (1970: 228-233), (28e)-(28g) are adapted from Safir (1992: 3, 26) and (28g) comes from Reinhart and Reuland (1993: 670). Similar naturally occurring examples were shown in (3) above.

restricted to *picture NP* reflexives: comparatives (e.g. *NPs like x-self, NP less/more AP than x-self*), exclusion phrases (e.g. *except/but/apart from x-self*) and coordinated NPs all appear to license it. The assumption that reflexives are never “discourse-bound” – i.e. covalued to a referent established in discourse, not necessarily within the same sentence – runs into trouble with the examples in (28i)-(28j). Note also that the reflexives in (28) could be replaced by pronouns without loss in acceptability. Insofar as the data in (28) contradict all of the fundamental assumptions in (27), they present serious problems for the CBT framework in its entirety.¹⁷

Belletti and Rizzi (1988) and Pesetsky (1995) proposed structures like (29) as a way to make some of the data in (28) compatible with a unified CBT. However, contrary to what they suggest, apparent violations of (27) do not gravitate around a single class of verbs with specific thematic properties, such as psych-verbs. The examples in (28b)-(28d) and (28f)-(28j) don’t involve psych-verbs at all.

¹⁷ The data here is restricted to English. Anaphoric systems of other languages – even closely related ones, such as Dutch and Frisian – present even deeper problems for the CBT. Dutch, for instance, has a tripartite pronominal system that does not fit Chomsky’s (1981) binary classification of anaphors *vs.* pronouns. The Dutch pronominal system consists of personal pronouns (e.g. *hem*), complex reflexive anaphors (e.g. *zichzelf*) and simplex expression (SE) anaphors (e.g. *zich*). The latter, as the examples below show, pattern with pronouns in some contexts and with complex reflexives in others (Reuland and Reinhart, 1995: 242):

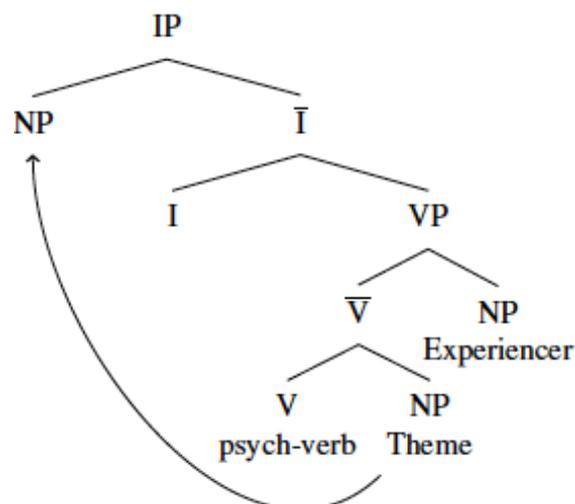
- (i) a. *Max wast zich / zichzelf / *hem*
 *Max washes SE / himself / *him*
 b. *Max haat zichzelf / *zich / *hem*
 *Max hates himself / *SE / *him*

Frisian, on the other hand, has a binary pronominal system like English, consisting of reflexive anaphors (e.g. *himsels*) and pronouns (e.g. *him*). Unlike English, however, pronouns are sometimes allowed to be locally free – contradicting a major prediction of the CBT (Reuland and Reinhart 1995: 243):

- (ii) *Max wasket him*
 Max washes him

Even more striking (and lexically unrestricted) examples of this are found in Traditional Jambi Malay (cf. Yanti et al. 2017). Obviously, neither these nor the facts in (28) constitute ultimate refutations or irremediable falsifications of the CBT, since no piece of empirical evidence has this kind of power. My point here is simply that there is no *independently motivated* way of tinkering with the CBT that could provide a natural account for all of these cases.

(29)



There is, however, an insight in proposals like (29), which reveals an important property of the reflexives in (28). It seems to be true that entities that stand in an Experiencer-like relation to propositions are, in a sense, favored antecedents for reflexives which contradict the assumptions in (27). This is clear in Pollard and Sag’s (1992: 277) contrast below:

- (30) a. Pictures of *himself* bothered [*John*]_{Experiencer}}
 b. *Pictures of *himself* bothered [*John’s father*]_{Experiencer}}

Insofar as (29) places Experiencers “higher” than Themes, (30) can be accounted within the CBT in terms of c-command. But this explanation does not extend to most cases in (28) — e.g. cases where the antecedent is either not a binder (i.e. does not c-command the reflexive) (e.g. (28c)), not an Experiencer (e.g. (28b)), or not local (e.g. (28d)-(28g)).

Furthermore, it does not cover paradigms like (31), due to Pollard and Sag (1992: 274), which are in some sense analogous to (30):

- (31) a. *John* was going to get even with Mary. That picture of *himself* in the paper would really annoy her, as would the other stunts *he* had planned.
 b. *Mary was quite taken aback by the publicity *John* was receiving. That picture of *himself* in the paper had really annoyed her, and there was not much she could do about it.

The antecedent in (31a) is not an Experiencer in the sense defined within the theta-theory. Moreover, it isn't even in the same sentence as the reflexive it should bind. The notion of a syntactically represented Experiencer is, thus, too narrow to offer a proper generalization for all of these cases.

What has generally been assumed is that the reflexives in (28) and (30)-(31) are conditioned by the *perspective* or *empathy* assumed in the discourse (Kuno 1987; Pollard and Sag 1992; Oshima 2007; Charnavel 2019). This is why they are called, somewhat loosely, *logophoric reflexives* (Reinhart and Reuland 1993), in analogy to the logophoric pronouns of African languages (cf. Hagège 1974). To be (only slightly) more precise, these reflexives must refer to entities whose viewpoints are represented or “empathized” within their discourse context.

Experiencers are, *ceteris paribus*, good candidates for being perspective bearers (i.e. logophoric antecedents). This explains the contrast in (30) without the need to invoke structures like (29). Since sentences with Experiencers are not the *only* kinds of sentences which license logophoric reflexives, proposing covert structures designed to make the odd binding properties of psych-verbs compatible with the CBT (e.g. (29)) address, at best, only a subset of a larger problem.

In sum, neither of the two strategies mentioned in connection to Lees and Klima (1963) offers a general enough way to save a unified CBT from these data: one cannot “get around” the counterexamples by positing covert structures or by enriching the rule system (e.g. by refining the notion of domain). Data like (28) and (30)-(31) suggest that *some* instances of reflexives can be licensed by discourse properties, rather than by syntactic conditions. This requires a revision of the core assumptions of the CBT. I will now turn to this issue and sketch my own attempt to reconcile a unified theory of reflexives with the diversity of data examined so far.

3 UNIVERSALITY AND VIOLABILITY: AN ALTERNATIVE ACCOUNT

At this, point, neither of the two strategies mentioned in the end of Section 1 (Enrich Structure and Enrich Principle) seem promising to handle data like (28). After the late 1980s, the standard response to exceptions to unified principles was to partition the domain of reflexives roughly between those to which rigid grammatical constraints apply (e.g. *Mary loves herself*) and those which are subject to (often vaguely stated) discourse conditions (logophoric cases like (28)). Crucially, in order

to make the distinction between grammar-abiding and grammar-exempt reflexives seem principled, the conditions for exemption are usually taken to be syntactically determined (cf. Pollard and Sag 1992; Reinhart and Reuland 1993). In Reuland's (2011: 92) words: "whether or not [a reflexive] is exempt is purely determined by its structural position". This is what I call the Two Reflexives Hypothesis (TRH).

My goal here is not to criticize the TRH — which I have done elsewhere (cf. Varaschin 2019)¹⁸ — but to suggest a *unified* alternative to the TRH that does not inherit the woes of Lees and Klima's (1963) simple theory nor of the CBT. This can be done by adopting a very simple general condition on reflexives and allowing it to be *violated* in certain contexts. The condition I propose states basically that reflexive pronouns impose a reflexive interpretation on their predicates. The universality of this condition accounts for the *unity* of reflexives (i.e. for the fact that logophoric and grammatical reflexives are identical in form) and its violability accounts for the fact that reflexives behave differently in different contexts. By *universality*, I mean the property of applying to potentially all reflexives in the language (without defining conditions for exemption within the grammar). By *violability* I mean that reflexives may fail to instantiate the terms of the condition and still be acceptable. Following Menuzzi (2004), therefore, the proposal sketched here will not treat logophoric reflexives like (28) as *exempt* from grammatical constraints, but as cases where a constraint is tolerably violated. The contexts which make such violations acceptable will also be specified below.

The general formalism I employ to state the grammatical constraint on reflexives will be Culicover and Jackendoff's (2005) theory of Simpler Syntax (SiSy) — in particular, its constructional rendition proposed in Culicover (2009, to appear). The theory is constraint-based, which means that its basic statements (i.e. constraints) amount to axioms in a formal logic with a model-theoretic interpretation. The models of the constraints are the expressions licensed by the

¹⁸ The main problem with the TRH is that, upon closer inspection, it turns out to be hard to isolate the exact syntactic conditions under which a reflexive can be exempt (cf. Dalrymple 1993; Menuzzi 2004; Charnavel 2019; Varaschin 2019). Besides that, the TRH ultimately entails a dubious claim to lexical homophony: if words are individuated not only by their phonology, but also in terms of their syntactic and semantic properties, grammatical reflexives and exempt reflexives count as distinct (albeit related) lexical items. As Charnavel (2019) argues, this suggestion seems particularly problematic in light of the fact that the formal coincidence between markers of reflexivity and logophoricity is not parochial to English, but is found in typologically unrelated languages: e.g. Icelandic (Maling 1984), Japanese (Kuno 1987), Turkish (Major and Ozkan 2018) and French (Charnavel 2019). Proponents of the TRH would have to stipulate the same unexplained homophony for each of these languages separately.

grammar (cf. Pullum and Scholz 2001). In SiSy, grammatical constraints are *constructions*, which are represented as curriesque signs (Curry 1963; Culicover to appear): i.e. tuples of parallel phonological, syntactic and semantic representations. The expressions they license — i.e. the expressions of which their partial descriptions are *true* — are called *constructs*.

SiSy also includes an independent tier for Grammatical Functions (GFs), which represents the syntactic argument structure of predicates. The basic units of the GF tier are Preds (short for *syntactic predicates*), which are composed of a sequence of ranked positions. These positions are not explicitly labeled as Subject or Object: such notions can be relationally defined as *first* GF, *second* GF, etc. Only syntactic arguments — i.e. the governable grammatical functions of LFG (cf. Dalrymple 2001: chap 1) — correspond to GFs on the GF tier. The main function of the GF tier is to aid in the mapping from semantic structures (SEM) to phrase structure (SYN). The ranking of GFs is determined according to some kind of accessibility hierarchy, in the sense of Kennan and Comrie (1977).

Since the GF tier will be important for my formulation of the constraint on reflexives, two examples of independent constructions involving the GF tier are given below. The format for representing constructions is an attribute-value matrix (AVM) — where natural numbers indicate correspondences between the different structures:

$$(32) \quad \mathbf{The\ Subject\ Construction:} \quad \left[\begin{array}{l} \text{SYN} \quad [s \text{ NP}_1 \text{ VP}]_2 \\ \text{GF} \quad [_{\text{Pred}} \text{ GF}_1 (\prec \dots)]_2 \end{array} \right]$$

$$(33) \quad \mathbf{The\ Passive\ Construction:} \quad \left[\begin{array}{l} \text{SYN} \quad [\dots V[\text{passive}] ([_{\text{PP}} P[\text{by}] [_{\text{NP}}]_1)]_2 \\ \text{GF} \quad [_{\text{Pred}} \text{ GF}_1 \prec [_{\text{Pred}} \text{ GF}_2]] \end{array} \right]$$

In English, (32) is the construction responsible for mapping Subjects (i.e. highest ranked GFs) to the sister of VP in SYN. (33) is the construction responsible for licensing passives. It expresses the intuition that the highest GF (the Subject) is “demoted” to an optional *by*-phrase, while the second-ranked GF (the Object) acquires status of the typical Subject of its Pred (i.e. it is adjacent to a left bracket). When (33) is unified with (32), this will entail that the second GF will get realized as the sister of VP. Note that (33) does not imply movement: it is just a truth-eligible declarative axiom.

We are now in a position to state, using the same format as above, the grammatical constraint responsible for licensing grammatical reflexives in English:

(34) The English Reflexive Construction (ERC):	$\begin{array}{ll} \text{PHON} & \Phi_{\text{English}}(\text{reflexive}, \varphi)_1 \\ \text{SYN} & \text{NP}[\text{reflexive}, \varphi]_1 \\ \text{GF} & [\text{Pred} \dots \text{GF} \dots \prec \text{GF}_1 \dots]_2 \\ \text{SEM} & \lambda x[\mathbf{2}' \dots (x) \dots (x) \dots] \end{array}$
---	--

Following the Paradigm Function Morphology of Stump (2001), the phonological part of the construction is the value of the paradigm function Φ_{English} applied to the lexical head and to the set of its φ -features (number, person and gender). For instance, $\Phi_{\text{English}}(\text{reflexive}, \{\text{'sing'}, \text{'3'}, \text{'fem'}\}) = /h3r'self/$. The GF tier encodes the information that the GF that corresponds to the reflexive must be the argument of a syntactic predicate with a higher ranked GF. This entails that there are no reflexive subjects in English (e.g. **Heself arrived*). The correspondence to the SEM tier requires, furthermore, that the syntactic predicate (Pred) of which the reflexive is an argument be *semantically* reflexive – i.e. that $\mathbf{2}'$ (i.e. the SEM counterpart of Pred) must have two arguments which are bound by the same λ -operator. This idea is borrowed from Keenan's (1988) and Szabolcsi's (1989) treatment of reflexives as reflexivizers (which correspond to duplicators in combinatory logic (cf. Steedman 1988)).

Note, furthermore, that (34) is a purely *lexical* (and, thus, language-specific) constraint. Since this constraint is, basically, a statement of the meaning of English reflexives as reflexivizers of their Preds, there is no need to interpret ERC as part of a dedicated module of UG. Furthermore, for this same reason, the violation of ERC in logophoric readings like (28) will be no more exotic than coercions and other kinds of standardized non-literality (cf. Levinson 2000), all of which can be seen as “violations” of the lexically encoded meanings of expressions.

Consider first the following concrete example of a simple reflexive construct which is licensed by the ERC. The PHON tier is simplified as English orthography:

(35)	$\begin{array}{ll} \text{PHON} & \text{Mary}_3 \text{ loves}_2 \text{ herself}_1 \\ \text{SYN} & [\text{s NP}[\text{Mary}]_3 [\text{VP V}[\text{loves}]_2 \text{ NP}[\text{reflexive}, \varphi]_1]] \\ \text{GF} & [\text{Pred GF}_3 \prec \text{GF}_1]_2 \\ \text{SEM} & \lambda P[P(\text{mary})]_3 (\lambda x[\text{love}'(x)(x)]_2) \end{array}$
------	--

The reflexive contained within construct in (35) is licensed because all of the relevant parts of the ERC in (34) are instantiated in (35). In PHON, *herself* is the value of Φ_{English} for a reflexive whose φ -features are {'sing', '3', 'fem'}. In GF, the reflexive's GF is an argument of a syntactic predicate which has a higher ranked GF. This syntactic predicate, moreover, corresponds to a reflexive predicate in SEM. In virtue of other constructions of the language (e.g. the interpretive rule for non-pronominal NPs and subject-predicate configurations) the highest GF corresponds, in SEM, to a generalized quantifier which combines with the reflexive predicate to yield (after the β -reduction steps) **love'**(mary)(mary). This is the correct semantics for the sentence — i.e. the sentence is true *iff* the pair <mary, mary> belongs to the set defined by **love'**.

Cases like (12a) above — i.e. *Mary's father supported herself* — are ruled out, not by ERC directly, but by *type-driven compositionality*: i.e. by the principle that ensures that the SEM of any branching SYN node γ is the result of applying the SEM of one of γ 's daughter nodes to the others. This means that, in (12a), the SEM of *Mary* cannot “reach out” of the NP headed by *father* to combine with the SEM of the predicate headed by *support*. The compositional rules of the language (which are also treated as constructions in SiSy) stipulate that the SEM of *Mary* must combine with the SEM of *father* before it combines with **support'** (see Reuland 2011: chap. 2).

More complex examples of constructs licensed by the ERC are given below. Each of these interact with *other* constructions: the imperative in (36), “raising” to subject in (37), “raising” to object in (38) and subject control in (39).¹⁹

$$(36) \quad \left[\begin{array}{l} \text{PHON} \quad \text{Shave}_2 \text{ yourself}_1 \\ \text{SYN} \quad [\text{s}, \text{VP} \text{ V}[\text{shave, imperative}]_2 \text{ NP}[\text{reflexive}, \varphi]_1] \\ \text{GF} \quad [\text{Pred} \text{ GF}_3 \prec \text{GF}_1]_2 \\ \text{SEM} \quad \lambda P[P(\text{you})]_3(\lambda x[\text{shave}'(x)(x)]_2) \end{array} \right]$$

$$(37) \quad \left[\begin{array}{l} \text{PHON} \quad \text{Mary}_3 \text{ seems}_4 \text{ to admire}_2 \text{ herself}_1 \\ \text{SYN} \quad [\text{s} \text{ NP}[\text{Mary}]_3 \text{ V}[\text{seems}]_4 [\text{VP} \text{ V}[\text{to admire}]_2 \text{ NP}[\text{reflexive}, \varphi]_1]] \\ \text{GF} \quad [\text{Pred} [\text{Pred} \text{ GF}_3]_4 \prec \text{GF}_1]_2 \\ \text{SEM} \quad \lambda P[P(\text{mary})]_3(\lambda y.\text{seem}'(\lambda x[\text{admire}'(x)(x)]_2(y)))_4 \end{array} \right]$$

¹⁹ Due to space limitations, I will not argue in favor of the particular analyses for these independent constructions. See Culicover (2009) for a relevant summary. I depart from prior analyses within SiSy by assuming a standard Montagovian PTQ-style semantics (Montague 1974) for the SEM tier.

- (38)
$$\left[\begin{array}{l} \text{PHON} \quad \text{Mary}_3 \text{ expects}_2 \text{ herself}_1 \text{ to win}_4 \\ \text{SYN} \quad [\text{S NP}[\text{Mary}]_3 [\text{VP V}[\text{expects}]_2 \text{ NP}[\text{reflexive}, \varphi]_1 [\text{VP V}[\text{to win}]_4]] \\ \text{GF} \quad [\text{Pred GF}_3 \prec [\text{Pred GF}_1]_4]_2 \\ \text{SEM} \quad \lambda P[P(\text{mary})]_3 (\lambda x[\text{expect}'(\lambda y[\text{win}'(y)]_4(x))(x))]_2 \end{array} \right]$$
- (39)
$$\left[\begin{array}{l} \text{PHON} \quad \text{Mary}_3 \text{ tries}_4 \text{ to like}_2 \text{ herself}_1 \\ \text{SYN} \quad [\text{S NP}[\text{Mary}]_3 [\text{VP V}[\text{tries}]_4 [\text{VP V}[\text{to like}]_2 \text{ NP}[\text{reflexive}, \varphi]_1]] \\ \text{GF} \quad [\text{Pred GF}_3]_4 [\text{Pred GF} \prec \text{GF}_1]_2 \\ \text{SEM} \quad \lambda P[P(\text{mary})]_3 (\lambda y.\text{try}'(\lambda x[\text{like}'(x)(x)]_2(y))(y))_4 \end{array} \right]$$

In each of these examples, the ERC licenses a reflexive without requiring devices such as empty categories or A-movement operations in SYN. Most of this work is turned over to the GF tier in SiSy. For instance, the fact that the highest ranked GF in (36) does not correspond to anything in SYN is part of the imperative construction. Likewise, the fact that *Mary* in (37) is, at once, the highest ranked syntactic argument of *seems* and of *like* is determined by the raising construction, which is formalized in terms of structure-sharing on the GF-tier. Structure sharing also occurs between the second GF of the predicate *expect* and the first GF of *win* in the raising-to-object structure (38). In (39), there is no structure sharing, but the reflexive, nonetheless, imposes a reflexive interpretation on the Pred which corresponds to *like*. The fact that the identical variables of the predicate **like'** ultimately (after all the β -reductions) get bound by the generalized quantifier $\lambda P[P(\text{mary})]$ is determined by the particular SEM associated with subject control.

What is important is that, in all of the constructs in (35)-(39), the reflexive is an argument of a syntactic predicate which maps into a reflexive predicate in SEM. All of these structures are, thus, models of a grammar which includes the ERC in (34): i.e. they satisfy the conditions imposed by this construction. This is crucially *not* the case for logophoric reflexives like (28), partially repeated (along with some of the examples in (3)) and reorganized below as (40)-(41).

- (40) a. *Max* boasted that the queen invited *Lucie* and *herself* for a drink.
 b. As for myself, I won't be invited.
 c. "[...] it was *herself* that *she* hated, and not the Little Sister."

- (41) a. The picture of *himself* in Newsweek shattered the peace of mind that *John* had spent the last six months trying to restore.
- b. *Albert* was never hostile to laymen who couldn't understand what physicists like *himself* were trying to prove.
- c. "*Pears* [...] believes some customers would be similar to *herself*."
- d. *John* believes that Mary would never consider marrying a man less wealthy than *himself*.

In none of these cases are the reflexives arguments of a syntactic predicate which corresponds to a reflexive predicate in SEM. The reflexives in (40) simply do not correspond to GFs: in (40a) the reflexive is *part* of a constituent which maps to a GF (the NP *Lucie and himself*), but not a bearer of a GF itself; in (40b-c), on the other hand, the reflexives occupy what MGG calls A'-positions (Topic and Focus, respectively). These are not assigned a GF in the theory of SiSy. The reason for this is that these positions are not possible targets for processes that motivate the GF-tier: e.g. syntactic alternations (e.g. passives) and structural case marking.

The problem with (41) is more telling. I assume that the reflexives in these examples *do* correspond to GFs — i.e., they *are arguments of syntactic predicates*. Nonetheless, they violate ERC because their syntactic predicates are not interpreted as reflexive in SEM. It seems, moreover, that the syntactic predicates therein are predicates for which *a reflexive interpretation is either infelicitous or undefined* (cf. Varaschin (2019) and Menuzzi (2004) for a thorough demonstration of this).²⁰

What both of these scenarios have in common is that they are *situations where ERC has no way of being fulfilled*. That is, in cases where a reflexive does not correspond to a GF as well as in cases where its GF is within a Pred that cannot be

²⁰ This is clearest for contrastive predicates such as the ones that appear in (41b-d). For these cases, Safir (1992) shows that reflexive interpretations would be either tautological (as (ia-b)) or contradictory (as (ic)). In either case, the result is uninformative, and, hence, infelicitous.

- (i) a. #*The physicists* are (very much) like *themselves*.
 b. # *Pears* is similar to *herself*.
 c. #*John* is less wealthy than *himself*.

In picture NPs like (41a), the Pred corresponds to a representational relation — in that case, to $\lambda x.\lambda y.$ **picture'**(x)(y). I argue in Varaschin (2019) that representational relations cannot be reflexive: if *x* is a representation of *y*, *x* and *y* are necessarily distinct. This seems intuitively obvious, but it also follows from a semantics of representational predicates wherein represented entities and their representations actually correspond to different semantic types. Imposing a semantic reflexivity on a representational predicate — as would be required by ERC — would, thus, amount to an *undefined* interpretation.

reflexive in SEM, violating ERC is, in some sense, *inevitable*. My hypothesis is that this inevitability is what distinguishes acceptable violations of ERC in logophoric readings from genuinely ungrammatical violations of ERC like (2), repeated below as (42):

- (42) a. *An old boyfriend of *Janet* still loves *herself*.
 b. **Janet* told me that Brad loves *herself*.
 c. *Janet's mom loves themselves.

In (42), the reflexives are in contexts where ERC *could have been satisfied*: they are syntactic arguments of their predicates (i.e. they bear GFs) and the semantic predicate **love'** is one for which a reflexive interpretation would be felicitous.

The idea, then, is that logophoric readings emerge in the context of *inevitable violations of ERC*. Since ERC is the *only* construction specifying the semantic contribution of reflexives in English (and this is what makes the account unified), when ERC cannot apply, reflexives must be interpreted *non-literally*, by means of a pragmatic procedure. This is expressed in what I call the Logophoric Strategy:

- (43) **Logophoric Strategy (LS)**: When a reflexive cannot be the argument of a syntactic predicate which is semantically reflexive, it is interpreted as a free variable whose value is determined according to general principles of discourse anaphora.

In sum: ERC states that the lexical duty of reflexives is to indicate that their Preds are reflexive in SEM. When they cannot do so (either because they do not correspond to GFs or because their Preds *can't* be reflexive in SEM), they must be interpreted by means of some other strategy. This is where logophoric readings kick in.

One can certainly be much more specific about *why* logophoric readings emerge precisely in contexts where ERC has no way of being fulfilled. According to Menuzzi (2004), violations of the condition on reflexives are acceptable in these contexts because they yield logophoric interpretations as a kind of generalized

implicature (Grice 1989). This seems to be a promising approach, however, due to space limitations, I will not explore it further in this paper ²¹

Another important observation is that the LS in (43) is only predicted to exist in languages that have something like the ERC in (34). In languages that lack a lexical form which encodes a reflexivizing function, there is nothing to be violated in logophoric contexts. Therefore, my prediction for these kinds of languages is that there are no logophoric reflexives like there are in English. Likewise, if a language has *lexicalized* logophoric interpretations for specific items— as seems to be the case with the Japanese reflexive *zibun* (Oshima 2007)— there should be no requirement that logophoric readings emerge only in the circumstances specified in the LS.

The main point I wanted to argue for is that violable constraints provide the flexibility to reconcile the variability of English reflexives with the uniformity of principles underlying them. By keeping the condition on reflexives simple and allowing it to be violated in certain contexts, we can avoid the main problems which jeopardized previous unified approaches such as Lees and Klima's and the CBT.

The particular condition I proposed — the ERC — captures the fact that there are “core cases” of reflexives which are defined by the grammar and logophoric cases, whose acceptability is explained by functional or pragmatic factors (e.g. the LS). As Menuzzi (2004) notes, this approach also explains why logophoric reflexives are perceived as “marked” outside of context. Another corollary of this, which seems to be empirically correct, is that the acceptability of logophoric reflexives is more subject to inter-speaker variation than that of reflexives which simply conform to the ERC.

FINAL REMARKS

In this paper, I presented a historical survey of unified approaches to reflexives within Mainstream Generative Grammar (MGG). I focused particularly in two prominent versions of this approach: the transformational account of reflexivization

²¹ The general idea behind a Gricean account of logophoric reflexives this: violations of grammatical conditions can be interpreted as violations of the Maxim of Manner, which governs *how what is said is to be said*. When the speaker violates ERC in logophoric readings, he is, therefore, violating the Maxim of Manner by using the reflexive inappropriately. This violation is pointless unless the speaker intends to communicate something by the reflexive *other than its literal function*, which is to signal semantic reflexivity (as dictated by ERC). In cases where the literal function of the reflexive *cannot apply* (i.e. in cases where ERC cannot be fulfilled), this intention to communicate something else by the reflexive (e.g. emphasis, empathy or perspective) is always evident. Therefore, the hearer assumes that this is what the speaker has in mind and derives a logophoric reading as an implicature (cf. Menuzzi 2004).

in Lees and Klima (1963) and the Classical Binding Theory (CBT) of Chomsky (1981, 1986) Huang (1983) and Reinhart (1983). Each of these is representative of a particular strategy in handling potential objections to unified principles: Lees and Klima's theory opts for the enriching structure and the CBT almost always opts for making the principles responsible for the interpretation of reflexives more complex.

In Section 3, I suggested a third strategy to preserve unified approaches from counterexamples: instead of enriching structure or introducing complexity into the principles (which usually entails introducing complexity into UG), we can keep both the principles and the structures simple if we allow the conditions on reflexives to be *violated in certain contexts*. These violations can, in turn, explain how logophoric readings are pragmatically derived by means of a non-literal interpretive strategy.

Though many technical details remain open,²² I proposed that the principle governing English reflexives is a *construction*. This entails that, at least in some respects, it is a *language-specific* pairing of form and meaning. It remains to be explored which aspects of ERC follow from general principles of human biological endowment and which ones are, in fact, parochial to individual languages.

It seems plausible, for instance, as suggested by Culicover (to appear), that the notion of a reflexive predicate in SEM is a universal property of the human conceptual system. The pressure to *mark* semantic reflexivity by means of some morphosyntactic cue also seems to be a universal — plausibly due to functional/pragmatic factors (cf. Haspelmath 2008). However, the way particular languages choose to mark these reflexive predicates is subject to great variation. The constructional approach, insofar as it reduces a general principle of grammar (the Condition A of Chomsky (1981)) to a lexical fact (i.e. to the ERC), turns out to be the right one in allowing for the full-range of cross-linguistic diversity found in anaphoric systems (cf. Dalrymple 1993). The idea that language-specific constructions are *violable* is a useful addition to the constructional stance, since it easily accounts for the diversity of reflexives *within an individual language* as well.

²² For example, I did not attempt to derive, within SiSy, the partial complementarity between reflexives and pronouns that is so important for unified approaches like Lees and Klima's and the CBT. Ideally, it would be desirable if this partial complementarity was not encoded as part of the construction which licenses pronouns, but would follow from independent pragmatic principles (e.g. from the fact that reflexives are *more informative/less ambiguous than pronouns*). See Levinson (2000: chap. 4) for richly articulated proposal along this lines.

ACKNOWLEDGMENTS

Special thanks to Peter Culicover for the constant dialogue and inspiration. Isabelle Charnavel, Ray Jackendoff, Richard Larson, Rafaela Miliorini and Heronides Moura provided very helpful comments on our conversations about these topics. Thanks also to Bob Levine, Ashwini Deo, Jordan Needle, Symon Stevens-Guille and audiences at Synners-OSU for the feedback. All errors are, of course, my own.

REFERENCES

- BELLETTI, Adriana; RIZZI, Luigi. Psych-verbs and theta-theory. **Natural Language and Linguistic Theory**, v. 6, n. 3, 1988.
- CANTRALL, William. **On the nature of the reflexive in English**. PhD Thesis. University of Illinois, Urbana-Champaign, 1969.
- CHARNAVEL, Isabelle. **Apparent exemption from Condition A**: a perspective-based theory. Unpublished Manuscript. 2019.
- CHOMSKY, Noam. **Syntactic structures**. The Hague: Mouton, 1957.
- CHOMSKY, Noam. **Aspects of the theory of syntax**. Cambridge: MIT Press, 1965.
- CHOMSKY, Noam. **Semantics in generative grammar**. The Hague:Mouton, 1972.
- CHOMSKY, Noam. Conditions on transformations. In: ANDERSON, Stephen; KIPARSKY, Paul (Eds). **A Festschrift for Morris Halle**. New York: Holt, Reinhart & Winston, 1973.
- CHOMSKY, Noam. **Lectures on government and binding**. Dordecht: Foris, 1981.
- CHOMSKY, Noam. **Knowledge of language**. New York: Praeger, 1986.
- CULICOVER, Peter W. **Natural language syntax**. Oxford: Oxford University Press, 2009.
- CULICOVER, Peter W. **Language universals, variation and change: a constructional approach**. New York: Oxford University Press, to appear.
- CULICOVER, Peter W.; JACKENDOFF, Ray. Something else for the binding theory. **Linguistic Inquiry**, v. 26, n. 2, 1995.
- CULICOVER, Peter W.; JACKENDOFF, Ray. **Simpler syntax**. Oxford: Oxford University Press, 2005.

CURRY, Haskell B. Some logical aspects of grammatical structure. In: JACOBSON, Roman (ed.). **Structure of language and its mathematical aspects: Proceedings of the twelfth symposium in applied mathematics**, s.l., 1963.

DALRYMPLE, Mary. **The syntax of anaphoric binding**. Stanford: CSLI, 1993.

DALRYMPLE, Mary. **Lexical-Functional Grammar**. New York: Academic Press, 2001.

EMONDS, Joseph. **Root and structure-preserving transformations**. PhD Thesis. Massachusetts Institute of Technology, Cambridge, 1970.

EVERAERT, Martin. **The syntax of reflexivization**. Dordrecht: Foris, 1986.

FIENGO, Robert; MAY, Robert. **Indices and identity**. Cambridge: MIT Press, 1994.

GRICE, Paul. **Studies in the Way of Words**. Cambridge: Harvard University Press, 1989.

HAGÈGE, Claude. Les pronoms logophoriques. **Bulletin de la Société de Linguistique**, v. 69, 1974.

HASPELMATH, Martin. A frequentist explanation of some universals of reflexive marking. **Linguistic Discovery**, v. 6, n. 1, 2008.

HEIM, Irene. Anaphora and semantic interpretation: a reinterpretation of Reinhart's approach. In: SAUERLAND, U; PERCUS, O. (eds.), **The Interpretative Tract: MIT Working Papers in Linguistics 25**. Cambridge: Department of Linguistics and Philosophy, MIT. 1998.

HORNSTEIN, Norbert. **A theory of syntax**. Cambridge: Cambridge University Press, 2009.

HUANG, C.T. A note on the binding theory. **Linguistic Inquiry**, v. 14, n. 3, 1983.

JACKENDOFF, Ray. Quantifiers in English. **Foundations of Language**, v. 4, 1968.

JACKENDOFF, Ray. **Some rules of semantic interpretation for English**. PhD Thesis, Massachusetts Institute of Technology, Cambridge, 1969.

JACKENDOFF, Ray. On Larson's treatment of the double object construction. **Linguistic Inquiry**, v. 21, n. 3, 1990.

JESPERSEN, Otto. **Essentials of English Grammar**. London: Edward Arnold, 1933.

KEENAN, Edward.; COMRIE, Bernard. Noun phrase accessibility and universal grammar. **Linguistic Inquiry**, v. 8, n. 1, 1977.

KEENAN, Edward. On semantics and the binding theory. In: HAWKINS, John. (Ed.) **Explaining Language Universals**. Oxford: Blackwell, 1988.

KOOPMAN, Hilda; SPORTICHE, Dominique. The position of subjects. **Lingua**, v. 85, 1991.

LANGACKER, Ronald. On pronominalization and the chain of command. In: REIBEL, D. A.; SCHANE, S. A. (Eds.). **Modern Studies in English**. Prentice Hall, 1969.

LARSON, Richard. Double objects revisited: reply to Jackendoff. **Linguistic Inquiry**, v. 21, n. 4, 1990.

LEBEAUX, David. **Locality and anaphoric binding**. The Linguistic Review, v. 4, n. 4, 1985.

LEES, Robert B.; KLIMA, Edward S. Rules for english pronominalization. **Language**, v. 39, n. 1, 1963.

LEVINSON, Stephen. **Presumptive meanings: the theory of generalized conversational implicature**. Cambridge: MIT Press, 2000.

LIDZ, Jeffrey; IDSARDI, William J. Chains and phonological form. **University of Pennsylvania Working Papers in Linguistics**, ERIC, v. 5, 1998.

MAJOR, Travis; OZKAN, Sozen. Anaphora in two turkic languages: Condition A is not enough. **IULC Working Papers**, v. 18, n. 2, 2018.

MALING, Joan. Non-clause-bounded reflexives in modern Icelandic. **Linguistics and Philosophy**, v. 7, n. 3, 1984.

MENUZZI, Sergio. Non-conflicting violations of grammatical constraints?. **Ilha do Desterro**, v. 1, n. 47, 2004.

MONTAGUE, Richard. **Formal philosophy**. New Haven: Yale University Press, 1974.

OSHIMA, David. On empathic and logophoric binding. **Research on language and computation**, v. 5, n. 1, 2007.

PESETSKY, David. **Zero syntax**. Cambridge: MIT Press, 1995.

PETERS, Stanley P.; RITCHIE, R.W. On the generative power of transformational grammars. **Information Sciences**, v. 6, 1973.

POLLARD, Carl.; SAG, Ivan. Anaphors in English and the scope of binding theory. **Linguistic Inquiry**, v. 23, n. 2, 1992.

POSTAL, Paul. **Cross-Over Phenomena**. New York: Holt, Rinehart and Winston, 1971.

PULLUM, Geoffrey; SCHOLZ, Barbara. On the distinction between generative-enumerative and model-theoretic syntactic frameworks. In: DE GROOTE, P.; MORRILL, G.; RETOR, C. (eds) **Logical Aspects of Computational Linguistics: 4th international conference**. Berlin: Springer, 2001.

REINHART, Tanya. **The syntactic domain of anaphora**. PhD Thesis, Massachusetts Institute of Technology, Cambridge, 1976.

REINHART, Tanya. **Anaphora and Semantic Interpretation**. Chicago: University of Chicago Press, 1983.

REINHART, Tanya. **Interface strategies: optimal and costly computations**. Cambridge: MIT Press, 2006.

REINHART, Tanya; REULAND, Eric. Reflexivity. **Linguistic Inquiry**, v. 24, n. 4, 1993.

REULAND, Eric; REINHART, Tanya. 1995. Pronouns, anaphors and case. In: HAIDER, Hubert; OLSEN, Susan; VIKNER, Sten (Eds.), **Studies in comparative Germanic syntax**. Dordrecht: The Netherlands: Kluwer, 1995.

REULAND, Eric. **Anaphora and Language Design**. Cambridge: MIT Press, 2011.

REULAND, Eric. Reflexives. In: THOMASON, S. (Ed.). **Oxford Handbooks online**. [S.l.]: Oxford University Press, 2014.

ROSENBAUM, Peter S. **The grammar of English predicate complement constructions**. PhD Thesis, Massachusetts Institute of Technology, Cambridge, 1967.

ROSS, John R. On declarative sentences. In: JACOBS, R. A.; ROSENBAUM, P. S. (Ed.). **Readings in English transformational grammar**. [S.l.]: Ginn-Blaisdell, 1970.

SAFIR, Ken. Implied non-coreference and the pattern of anaphora. **Linguistics and Philosophy**, v. 15, n. 1, 1992.

SAFIR, Ken. Anaphors, movement and coconstrual. In: DELSING, L-O.; et al. (Ed.). **Grammatik i Fokus. Grammar in focus: A festschrift for Christer Platzack**. Lund: Lund University, Department of Scandinavian Languages, 2003.

SAFIR, Ken. **The Syntax of Anaphora**. Oxford: Oxford University Press, 2004.

STEEDMAN, Mark. Combinators and grammars. In: OHERLE, R.; BACH, E.; WHEELER, D. (eds.) **Categorial grammars and natural language structures**. Dordrecht: Foris, 1988.

STUMP, Gregory. **Inflectional morphology: a theory of paradigm structure**. Cambridge: Cambridge University Press, 2001.

SZABOLSCI, Anna. Bound variables in syntax (are there any?). In: BARTSCH, R.; VAN BENTHAM, J.; BOAS, P. (eds.) **Semantics and contextual expression**. Dordrecht: Foris, 1989.

VARASCHIN, Giuseppe. Logophoricity and Antireflexivity. **Unpublished manuscript**. 2019

WILLIAMS, Edwin. PRO and subject of NP. **Natural Language and Linguistic Theory**, v. 3, n. 3, 1985.

YANTI et al. The grammar of binding in the languages of the world: A response to Reuland. **Cognition**, v.168, n.1, 2017.

ZRIBI-HERTZ, Anne. Anaphor binding and narrative point of view: English reflexive pronouns in sentence and discourse. **Language**, v. 65, n. 4, 1989.