Chapter 2

Theoretical Background (Pro-Drop)

2.0. Introduction

The goals of this chapter are (1) to provide a critical review of pro-drop as a

theoretical construct that has commonly been used to explain and classify differences

between languages, (2) to review how recent research in Optimality Theory (Prince and

Smolensky 1993) challenges previous conceptions of a 'pro-drop parameter' and

demonstrates the need to consider interactions of constraints from various levels of

grammar, and (3) to address recent critiques that these Optimality Theoretic approaches

violate the principle of the autonomy of syntax.[1]

To achieve these goals, Section 2.1 begins with a discussion of the cluster of

properties that have been traditionally subsumed under the pro-drop designation. Section

2.2 discusses some of the early proposals to account for the cross-linguistic distribution

of null subjects through rule-based application of universal principles of grammar.

Section 2.3 reviews the move within syntactic theory to situate the pro-drop construct

within a parameter-setting framework (Chomsky 1981); this section also discusses the

major pro-drop proposals that emerged from that framework. Section 2.4 considers later

proposals regarding pro-drop, focusing on theories of identification and morphological

uniformity. Section 2.5 then considers early Minimalist accounts of pro-drop during the

beginning of the last decade. Section 2.6 reviews a significant shift in the linguistic

theorizing that provided new tools for reconsidering the pro-drop question—Optimality Theory. This section provides a brief introduction to the major tenets of Optimality Theory (OT) and reviews the OT studies that pertain to interfaces between syntax and discourse, particularly as these interfaces relate to pro-drop.

Since this dissertation represents the first developmental account to characterize the acquisition of several features of pro-drop through the interaction of discoursal and syntactic constraints within an OT framework, affirming the legitimacy of such an approach is crucial. For this reason, Section 2.7 reviews, and responds to, one recent critique (Newmeyer, 2000) of the use of OT to address issues of syntax, particularly when these approaches use constraints from numerous grammatical levels. Finally, having discussed the major proposals regarding pro-drop to date and having defended an OT-syntax approach to the pro-drop question, Section 2.8 reviews some of the limitations of the earlier proposals, highlighting especially the need for empirical evidence from acquisition studies, and for developmental analyses that reflect that evidence.

2.1. Defining pro-drop

Linguistic theorizing during the last two decades has often attempted to explain why some languages permit subjects of tensed clauses to be null, and other languages do not. For example, certain utterances that are permitted in languages such as Portuguese (2.1a), Italian (2.1b), and Spanish (2.1c) are ungrammatical in a language such as English (2.2).

(2.1.) a. *Pedro disse que viu o João.* (Portuguese)
Pedro$_i$ say-3sg-pst that pro$_i$ saw-3sg-pst João
'Pedro said that he saw João.'

b. *Ø piove.* (Italian)
　　　　　　rain-3sg-prs
　　　　　'(It) is raining.'

　　　　c. *Ø Compramos　unos　recuerdos.* (Spanish)
　　　　　　buy-1pl-pst　some souvenirs
　　　　　'We bought some souvenirs.'

(2.2.)　　a. **Pedro said that saw Juan.*

　　　　b. **is raining.*

　　　　c. **bought some souvenirs.*

Since Chomsky (1981), it has often been thought that sentences such as those in 2.1 contain a phonetically empty, but structurally present, subject. Chomsky suggested that these tensed sentences contain an empty pronominal element, referred to as 'pro'. One difference between the languages in 2.1 and languages that do not that permit such utterances amounts to whether or not a language permits the use of this pro element in tensed clauses. Those that do permit its use have been labeled 'pro-drop' or 'null subject' languages, while those that do not have been labeled 'non-pro-drop' or 'non-null subject' languages.

Chomsky (1981) recognized that pro-drop languages typically share a number of characteristics that differentiate them from non-pro-drop languages. For example, as others had earlier noted (Jespersen 1924, Perlmutter 1971, Taraldsen 1978), many pro-drop languages have 'rich' inflectional systems. Chomsky argued that, since the recovery of the missing subject was critical, these 'richer' language systems had unique qualities related to agreement (AGR) and inflection (INFL) that permitted this recoverability.

In addition to rich agreement, pro-drop languages were thought to typically display a cluster of associated morphological and syntactic properties. The following

examples from the pro-drop languages of Italian (adapted from Chomsky 1981), and

Spanish (adapted from Liceras and Díaz 1999), illustrate this cluster of properties.

Examples 2.3a and 2.3b illustrate the prototypical omission of a subject pronoun in a

tensed clause:

(2.3.)   a. ∅ *ho*   *trovato*   *il libro.*   (Italian)
              b. ∅ *he*   *encontrado*   *el libro.*   (Spanish)
                have-1sg find-part-pst   the book
                'I have found the book.'

Pro-drop languages also often permit inversion in declarative sentences, as in 2.4,

where the subject follows the verb:

(2.4.)   a. ∅ *ha*   *mangiato*   *Giovanni.*   (Italian)
              b. ∅ *ha*   *comida*   *Juan*   (Spanish)
                have-3sg eat-part-pst John
                'John has eaten.'

In addition, pro-drop languages permit apparent violations of *that-trace*. This

refers to a rule that, at least in English, prohibits clauses containing both the

complementizer *that* and a 'trace', a phonetically null element left behind when a

syntactic element moves in a clause as, for example, in '*that t will leave*' in 2.5, where

extraction of the subject from the lower clause requires the absence the complementizer

in English, but not in Italian or Spanish:

(2.5.)   a. *Chi*$_i$ *credi*   *che t*$_i$ *partirà?*   (Italian)
              b. *¿Quién*$_i$ *crees*   *que t*$_i$ *se irá?*   (Spanish)
                who$_i$ think-2sg-prs that *t*$_i$ leave-3sg-fut
                'Who do you think (*that) will leave?'

Still other properties sometimes associated with pro-drop languages include the allowance of certain utterances that cannot easily be given English equivalents, for example, those containing long *wh*-movement (2.6) and empty resumptive pronouns (2.7) ('a' examples are Italian; 'b' examples, Spanish):

(2.6.) a. *L'uomo$_i$ che mi domando chi $\varnothing_i$ abbia   visto.*
        b. *El hombre$_i$ que me pregunto a quién $\varnothing_i$ había   visto.*
           the man   that I   wonder who   has-3sg-pst see-part-pst
           *'The man (x) such that I wonder who x saw.'

(2.7.) a. *Ecco   la ragazza$_i$ che mi domando chi   credi   che $\varnothing_i$ farla.*
        b. *Esta es la chica$_i$ que me pregunto quién cree   que $\varnothing_i$ lo hizo.*
           this   is the girl$_i$ that I wonder   who think-3sg that   did it
           *'This is the girl who I wonder who thinks that (she) did it.'

Expletive subjects (2.8) are often also included as a property of pro-drop languages:

(2.8.) a. *$\varnothing$ ha   piovuto.*   (Italian)
        b. *$\varnothing$ ha   llovido.*   (Spanish)
           $\varnothing$ has-3sg-pst rain-part-pst
           '(It) has rained.'

As will be shown later, not all these of these characteristics cluster together in every language that permits some of them, and this has led to some disagreement among researchers over what should or should not be included in a definition of pro-drop; nevertheless, there is broad acceptance that these languages contain empty categories and have phonetically unrealized, but syntactically present, elements. Furthermore, most have come to view pro-drop as a deep unlearned generalization, a universal, but parameterized, principle of grammar.[2]

2.2 Early proposals regarding pro-drop

Empty categories, those assumed to be syntactically present but phonetically null, have played a major role in syntactic theorizing. Early debate on pro-drop focused on distinguishing the characteristics of various empty categories, especially the empty category called 'little pro' (Chomsky 1982). Taraldsen (1978) had argued that null subjects are all empty Noun Phrases (NPs) bound in S' by subject-verb agreement, as in 2.9, with the missing subject permitted because of the 'rich' verbal inflectional system.[3]

(2.9.)   [$_{NPi}$ *e*] *comen*$_i$ *a las diez*.   (Spanish)
              eat-3PL at ten o'clock
              'They eat at ten o'clock.'

Jaeggli (1980), Chomsky (1981), and Suñer (1982) agreed with Taraldsen (1978) on the importance of a rich inflectional system, too, but they argued that the null element was not an empty NP but rather an empty element (PRO) posited by Government and Binding Theory (GB, Chomsky 1981) for control constructions where a non-finite verb had a null subject.

These early accounts of pro-drop were largely rule-based. For example, Jaeggli (1980) and Chomsky (1981) start from the position that sentences found in a pro-drop language (2.10a-c) share with non-pro-drop languages a rule of affix movement such as (2.11) that attaches INFL to the first verbal element at the level of Phonetic Form (PF).[4]

(2.10.) a. *Gianni telefonerà.*   'Gianni will call.'   (Italian)

          b. *Telefonerà.*   'He will call.'

          c. *Telefonerà Gianni.*   'Gianni will call.'

(2.11.) INFL V→V + INFL

The rule in 2.11 leaves no trace, and in non-pro-drop languages, this rule always applies in the phonological component of the grammar, producing the structure in 2.12, where NP is governed by INFL:

(2.12.) NP' INFL' VP

PRO does not appear in a structure such as 2.12 because the subject position is governed by INFL, and PRO cannot be governed. But in pro-drop languages, rule 2.11 can apply not only in the phonological component, but also, optionally, in the syntax. When this option is not used, a sentence such as 2.9a surfaces, with the structure of 2.12 and nominative Case assignment applying to the NP governed by INFL. When the option is used, and 2.11 applies in the syntax, then the structure becomes that of 2.13, where the subject NP is not governed by INFL and so must become filled by PRO.

(2.13.) NP' [$_{VP}$ V + INFL' ...]

According to Chomsky (1981), The 2.10c case adjoins the subject to the VP, leaving a spot to be filled by a dummy pronominal (e.g. *there* or *it* in English), but in pro-drop languages, the minimal dummy allowed is PRO, resulting in 2.14:

(2.14.) [PRO' [$_{VP}$ [$_{VP}$ t*elefon* + INFL'] *Gianni'*]]

The structure in 2.14 not only satisfies the requirement that PRO be ungoverned, but it also allows the postverbal subject to receive nominative Case due to its being governed by INFL, as posited by GB.

Soon after this initial proposal using 'big' PRO, Chomsky (1982) introduced the empty category of 'pro' as a pronominal anaphor, justified as a separate category for several reasons, including: (1) obligatory fronting of the verb in Spanish questions (cf. Torrego 1984) would place PRO in a position where it would be governed (and this is not possible); (2) the missing subject in tensed clauses cannot normally be arbitrary in reference (unlike PRO); and (3) positing a pro category would permit the elimination of the stipulation or parameter that the rule in 2.11 can apply in syntax in pro-drop languages.

GB thus distinguished a number of empty categories. For example, in 2.15 and 2.16, the 'emptiness' was thought to be due to movement (resulting in a trace):

(2.15.) *Mary$_i$ was kissed t$_i$ under the apple tree*.

(2.16.) *Who$_i$ do you think t$_i$ finished the pizza?*

Empty categories could also be distinguished in regards to their referential features: NP-traces were considered [+anaphoric] and [–pronominal], while variables (empty elements, such as *wh*-trace or traces left behind by quantifier raising, which must be A-bar bound by an operator) could be considered [-anaphoric] and [–pronominal].

In contrast, in 2.17 and 2.18, there was thought to be no movement; the empty element is the subject of a non-finite clause, a pronominal anaphor (PRO) that, in regard to GB's principles of Binding Theory, must be both bound and free within its governing category [+pronominal and +anaphoric]. Since this is not possible, PRO must remain ungoverned.

(2.17.) *Harold$_i$ wanted* PRO$_i$ *to see the doctor*.

(2.18.) *Sarah$_i$ hoped* PRO$_i$ *to find a new job.*

A final empty category was 'little pro' (introduced by Chomsky 1982), the category that is most significant to a discussion of pro-drop languages. This 'little pro' was the understood subject of finite verbs, as in 2.19 and 2.20. Unlike 'big PRO', this category has the full referential properties that a personal pronoun would have; it is a pronominal non-anaphor [+pronominal, -anaphoric], which is to say that under GB's Binding Theory, it is free in its local domain, not bound.

(2.19.)　pro *baila　　bien.*　　(Spanish, from Jaeggli 1982)
　　　　　dance-3sg　well
　　　　　'She dances well.'

(2.20.)　pro *verrà.*　　　　　(Italian, from Rizzi 1982)
　　　　　come-3sg-fut
　　　　　'He will come.'

As we shall see, GB theory treats sentences such as 2.19 and 2.20 as having an empty subject (different from no subject at all), because in this theory it is an inviolable principle that sentences have subjects (Extended Projection Principle, Chomsky 1981), regardless of whether or not they are visible.

Given the different distributions of these empty categories, a typology emerges Table 2.1:

Table 2.1 Typology of Empty Categories (from Jaeggli and Safir 1989:22)

	Pronominal	Anaphoric
PRO	+	+
pro	+	-
NP-trace	-	+
Variable	-	-

In addition to the typology given in Table 2.1, these empty categories could be contrasted in regards to how each category is handled by the different subtheories of grammar within the GB framework (e.g. Case theory, Theta-Theory, and Binding Theory).[5] For example, in a GB framework, Theta Theory provides the impetus for creating a structural position (even if not phonetically filled) to which a theta-role may be assigned. In the case of expletive elements that do not bear thematic roles, as is the case with the English *it* in 2.21, Theta Theory does not force an expletive element, so something else must.

(2.21.) (*It*) *seems Isabel is sick.*

This motivated the Extended Projection Principle (EPP, Chomsky 1981), which essentially states that every sentence must have a structural subject. The Spanish equivalent of 2.21, given in 2.22, does not overtly satisfy this principle, so the conclusion must be that EPP is satisfied covertly, by the existence of an empty category.

(2.22.) pro *parece   que Isabel está enfermo.*
         seems-3sg that Isabel is   sick
        'It seems that Isabel is sick.'

Thus, much of the early work in pro-drop revolved around identifying how the application of universal principles (together with the application of any rules) would account for the distribution of empty elements.

2.3 Parameter-setting approaches

A primary goal within linguistic theory has been to explain how children are able to acquire the complex nuances that a grammar presents them. Since it appears that children's linguistic competence goes beyond what the input they have received would make possible, innate knowledge of some principles of grammar has been assumed. In the conception of Chomsky's Principles and Parameters Theory (1981), children innately possess a set of universal and inviolable principles, as well as a set of parameterized options that can have more than one setting, and pro-drop has been thought to be one such parameter. If this parameter is set positively, the child's language will display the distinguishing characteristics of pro-drop languages. If set negatively, the child's language will not display these characteristics. Either way, the child's grammar must still adhere to the conditions or constraints set by the universal principles of grammar.

Early applications of the parameter-setting approach involved the creation of new principles and new parameters. For example, the analysis of Bouchard (1983) posited a 'Principle of Lexicalization' and introduced a new parameter regarding where nominals are assigned Case. Bouchard argued that nominals are lexical if and only if they contain person, number, gender, and Case at Phonetic Form (PF).[6] According to Bouchard, null subject languages consist of a parameter that optionally delays the percolation of Case assigning features of INFL until the level of Logical Form (LF). If the nominal is first assigned Case at LF by INFL, the nominal will not appear at PF. Bouchard then theorizes that in non-pro-drop languages the Case-assigning features of INFL must percolate up before surface structure, requiring Case-marked, overtly lexicalized nominals.

In another early paper, Suñer (1982) maintained that PRO and Ø are base generated and that *trace* is derived. She claims that traces are governed in both Spanish and English, but that PRO differs in these two languages.[7] According to Suñer, in English PRO is never governed or Case-marked, but in Spanish, government of PRO is correlated with [±] tense. In Spanish, PRO is ungoverned in nonfinite clauses 2.23, but governed in finite clauses 2.24.

(2.23.) *Paco quiere*   PRO *comer*.
         Paco wants-3sg   to eat
         'Paco wants to eat.'

(2.24.) PRO$_i$ *comen$_i$ a las diez*.
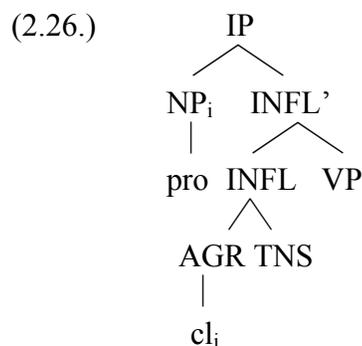   eat-3pl at   ten
         'They eat at ten o'clock'

The critical argument of Suñer (1982) is that PRO, pro, and traces in Spanish may appear in finite and nonfinite clauses, but in order to maintain this, she concludes that the government of PRO in tensed clauses in Spanish is a result of the lack of a subject requirement. For this, Suñer (1982) requires an additional parameter — the [±] obligatory subject parameter.

Rizzi (1982) argued that the *that-trace* effect is suspended in certain null subject languages because in these languages the subject may first postpose, adjoining to the VP, where it is properly governed by V, and then is extracted from this proposed position, represented in 2.25.

(2.25.) [$_{CP}$ *chi$_i$*] *credi*   [$_{S'}$ *che t$_i$ verrà t$_i$*]   (Italian)
             who believes that will come
         'Who$_i$ believes he$_i$ will come?'

While this idea requires a certain ordering of movement, it lends support to a claim that *that-trace* effects are not the result of a parameter, but rather a fortuitous outcome of a grammar where the possibility of inversion of subjects permits an otherwise illicit *that-trace* sequence.

Within the parameter-setting model, most generative accounts have addressed the pro-drop question by attempting to isolate (1) licensing conditions (i.e. what permits the appearance of the null subject in a sentence) and (2) identification processes (i.e. how the referent of the empty subject is semantically recovered. For example, Rizzi (1982) made the significant claim that null subject languages differ from non-null subject languages in that the former have verbal inflections that are specified with a [+pronoun] feature with clitic-like properties. The subject pronoun cliticizes onto the verb, permitting recovery of person and number features. This insight was pursued in subsequent works (e.g. Jaeggli 1982, Bouchard 1983, Rizzi 1986, Roberge 1990, et al.) where the structure was assumed to be such that subject clitics (cl) can be proper licensers of pro as in 2.26:

(2.26.)

```
              IP
            /    \
         NP_i    INFL'
          |      /    \
         pro  INFL    VP
               /  \
            AGR   TNS
              |
            cl_i
```

In 2.26, the subject clitic governs the [NP, IP] position. If lexical material is not present in the NP position, pro fills the position, receiving its licensing by the clitic under AGR. This provides an additional way for pro to be licensed. As Roberge (1990) observes, this proposal allows the licensing of pro in some languages (e.g. Spanish and

Italian) to occur through rich agreement, while licensing for pro in languages with subject clitics could be licensed by the clitics themselves. This divides languages into two additional categories — those with clitics and those without — and permits the parameterized option to be whether or not a language permits clitics. A general Recoverability Principle within a theory of government would still need to apply to both, meaning that empty categories must be licensed, but licensing could then come about in a variety of ways.

2.4 Later proposals regarding pro-drop

For many later studies in pro-drop, identification became the core property of null subjects. Early explanations of pro-drop were framed in terms of agreement features governing the empty category. The simple thought was that where there is overt agreement, subjects can be dropped. This observation alone frequently makes correct predictions. For example, Kenstowicz (1989) observed that person inflection was necessary for licensing in Bani-Hassan Arabic. In this dialect, if a verbal form inflects for tense, but not person (as in the participial form), then subject pronouns must be overt, but if it inflects for both person and tense, then null subjects are allowed (as in the perfect). Since it is possible that it is not a single parameter within a language that can blindly determine whether subjects can or cannot be dropped throughout a language, Kenstowicz (1989) provides evidence for a claim that more than just licensing is needed.

Rizzi (1986) also made such a claim when he posited that pro must be both licensed and identified. Earlier work in pro-drop had made much of the presence of verbal agreement features as the important element in licensing the empty subject

category. It had been assumed that when agreement features were rich enough, AGR possessed special governing powers. But Rizzi (1986) noticed that there are contexts where pro appears when identification through rich agreement is not possible. For example, in the Italian sentences in 2.27, where the objects are empty, the internal argument position requires pro, raising the problem of how this element is properly identified.

> (2.27.) a. *Questo conduce* [*la gente*/Ø] *alla seguente conclusione.*
>   'This   leads   (people)   to the following conclusion.'
>
>   b. *Gianni è sempre pronto ad accontentare* [*la gente*/Ø]
>   'Gianni is always ready   to please   (people).'

Rizzi's solution was to differentiate licensing from identification. Whereas licensing simply refers to the sanctioning of a given constituent, identification refers to the way the interpretation of empty categories is determined when only implicit null subjects are used. One language may choose INFL as a licenser for pro in the subject position and V for pro in the object position; another language may prohibit licensing altogether. Thus, the class of licensing heads may vary cross-linguistically, from permitting every head to be a possible licenser to allowing no licensing heads.

In terms of identification, Rizzi argued that referential pro is allowed if INFL is specified for the agreement feature of person, but the identification procedure is optional. Thus, some languages (e.g. German) will have an INFL specified for person, but will still not allow referential null subjects. Such languages will not be pro-drop. In other languages (e.g. Spanish) the null subject will have a grammatical specification of features on its INFL. Thus, pro will be both properly licensed and identified.

In contrast to Chomsky (1981) and Jaeggli (1982), for whom identification required agreement with the phi-features[8] in INFL (with licensing requiring Case and lack of government), in the proposal of Rizzi (1986), for identification to succeed, coindexation with rich agreement specification (or an extended notion of binding) was necessary with licensing involving government. Rizzi (1986:524) proposes the principle in 2.28:

(2.28.) pro is Case-marked by X_y where X is a governing head of the type y.

As noted, for Rizzi (1986), identification takes place through f-features or rich agreement specification. This is shown in 2.29, where pro resides in the Spec of INFL governed by INFL, and identified by the rich agreement specification on the first verbal form, *hemos*, in INFL.

(2.29.)　　INFL
　　　　　／￣＼
　　　Spec　　　I'
　　　　 |　　／￣＼
　　　pro_i　 I　　VP
　　　　　 |　　　|
　　　　he*mos*$_i$ encontrado
　　　　　have　　found

This analysis, still based on rich agreement features, partially works for languages such as Spanish and Italian, which have the option of a licensing head, but runs into difficulties with other languages such as Chinese or Korean, that lack agreement features but still permit null subjects.

The inadequacy of agreement alone, or even agreement and identification, to account for pro-drop cross-linguistically is taken up by Huang (1984, 1989) who demonstrated that not all pro-drop languages rely on overt agreement to recover the missing subjects. Huang (1984, 1989) demonstrates that Chinese, a language entirely lacking in agreement features, may drop subjects (and objects) from finite sentences, as in 2.30 and 2.31. According to Huang, the null arguments in 2.31 are optional.

(2.30.) *Zhangsan kanjian Lisi le   ma*?
Zhangsan see   Lisi Asp Q
'Did Zhangsan see Lisi?'

(2.31.) (*ta*) *kanjian* (*ta*) *le*.
he   see   him Perf
'(He) saw (him).

Since Chinese does not have overt agreement, the null subject in 2.31 is problematic for accounts that rely on local licensing and identification.[9] Huang (1989:193) addresses this problem through a generalized control theory outlined in 2.32 with the notion of control domain in 2.33 defined in Manzini (1983) and Nishigauchi (1984):

(2.32.) Generalized Control Rule (GCR)

An empty pronominal is controlled in its control domain (if it has one).

(2.33.) α is the control domain for β iff it is the minimal category that satisfies both (a) and (b):

a. α is the lowest S or NP that contains (i) β, or (ii) the minimal maximal category containing β.

b. α contains a SUBJECT accessible to β.

This rule specifies the environments in which pro and PRO (Huang treats both the same)

must have 'a local, unique, non-arbitrary antecedent' (1989:194). Crucially, if the empty

category does not have a control domain, then this type of control does not apply. In

such a case, reference may involve long-distance antecedents, arbitrary reference, or even

pragmatic considerations. Huang's (1989) position is summarized in Figure 2.1:
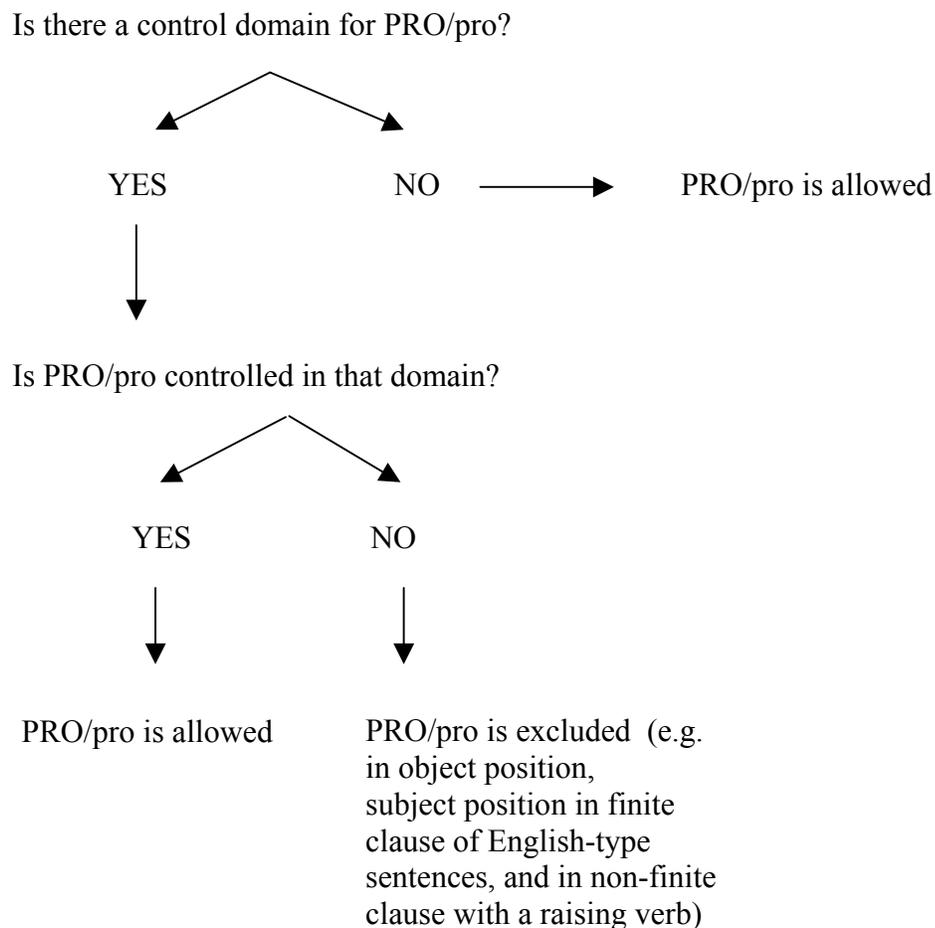
Figure 2.1. Distribution and reference of PRO and pro

Is there a control domain for PRO/pro?

YES   NO ⟶   PRO/pro is allowed

Is PRO/pro controlled in that domain?

YES   NO

PRO/pro is allowed   PRO/pro is excluded (e.g.
in object position,
subject position in finite
clause of English-type
sentences, and in non-finite
clause with a raising verb)

Figure 2.1 shows where PRO/pro is allowed according to Huang (1989). When it

is allowed, its reference is either free (when there is no controller) or determined (when it

has a control domain and is controlled in that domain). One important claim Huang

(1989) makes is that where PRO is excluded, so is pro. He is thus able to conflate the

two categories PRO and pro into a single category, eliminating the need for Chomsky's PRO Theorem (1981) that separately argued for an ungoverned PRO based on Binding Theory.

Jaeggli and Safir (1989) represent another analysis that attempted to account for the presence of pro-drop in both languages with rich inflectional systems and those lacking an inflectional system altogether. Jaeggli and Safir reformulated the pro-drop parameter with an argument that null subjects are permitted in *all and only* languages with morphologically uniform inflectional paradigms. Jaeggli and Safir essentially claim that languages lacking morphological uniformity will not allow null subjects. Their definition of 'morphological uniformity' is given in 2.34:

(2.34.) Morphological Uniformity
An inflectional paradigm P in a language L is morphologically uniform iff P has either only underived inflectional forms or only derived inflectional forms.

The Morphological Uniformity Hypothesis (MUH) essentially claims that languages that lack morphological uniformity will not permit null subjects. For example, English and French lack morphological uniformity and also do not permit null subjects. In contrast, Spanish and Italian do exhibit (relatively) uniform inflectional paradigms in the richness of their morphology so they permit pro-drop; Chinese and Korean display uniformity by virtue of their complete lack of verbal inflections, so they also permit pro-drop. Thus, the MUH would predict that the only pro-drop languages are those that have this quality of a uniform paradigm.

One final claim often made in the early 1980s was that null subjects and free inversion were linked by the same parameter, or the same syntactic requirement (Chao

1980, Jaeggli 1980, Chomsky 1981, et al.). Chao (1980) and Safir (1985) challenged this claim based on certain Italian dialects that permitted inversion, but not null subjects, and Portuguese, which permits null subjects but not inversion. Although the details of the analyses of Chao (1980) and Safir (1985) are not critical here,[10] the observation that inversion and the absence of null subjects are not necessarily conjoined points to the need for an analysis that treats these two phenomena as the result of conditions that could variably produce one or the other, or both, effects.

2.5 Minimalist accounts

During recent years, the rise of the Minimalist Program (MP, Chomsky 1995) as a dominant theory of UG has highlighted the core assumption that a radical economy in a grammar's set of theoretical and descriptive apparatus is necessary. To some extent, the MP and OT both share a certain set of insights that have gained greater acceptance. First, both MP and OT depart from classical GB theory by introducing the notion that different derivations, mappings, and surface forms compete with each other. Particularly in OT, but to a degree also MP, the candidate that wins is the one that is 'best' in relation to the other candidates. Secondly, both MP and OT recognize that the syntactic component cannot be entirely autonomous — it can only be understood with reference to the lexical morphosyntactic and semantic systems of the grammar. These developments in grammatical theory permit new consideration of questions that have been resistant to explanation when viewed only on the level of syntax. Pro-drop is one of those questions.

Speas (1994) considered pro-drop from a Minimalist stance. Using Chomsky's (1991) Principle of Economy ('project XP only if XP has content,'), Speas proposed that

what separates pro-drop from non-pro-drop languages is where affixes are generated. Some languages generate affixes in the syntax, others in the lexicon. If AGR features are attached to the verb, projection is not possible, resulting in a non-pro-drop language. If AGR features are not attached, that is, if there is a morpheme in AGR (or in the case of some languages, no AGR at all), then pro-drop becomes possible. In languages such as Chinese that do not have agreement morphology, there is no need for an AGRP projection, and 'no need' translates into no projection and no requirement for licensing conditions. For Speas (1994), the pro-drop parameter is a parameter based on whether inflection is syntactic or lexical.

In another Minimalist approach, Radford (1997) argues that languages differ in the strength of the agreement features carried by their finite verbs.[11] He suggests that when finite verbs carry strong agreement features, nonauxiliary finite verbs can raise from V to INFL and they can have a null pro subject, but when verbs carry only weak agreement features, neither the raising nor the null subject is possible. Radford assumes that the strength of features is correlated to the richness of the agreement inflections, and that, in a language that has a rich system of agreement inflections, identification of the null subject is more easily recoverable. As an illustrative example, Radford uses the historical changes to the inflectional system of Old English (which he claims was pro-drop) that eventually yielded Modern English (which is not pro-drop).

Minimalist theory employs 'feature checking', a relation between two elements such that one or more designated features they share are eliminated, as in 2.35:

(2.35.) a. *When did you go?*

b. **They went when.*

29

In 2.35a the +*wh* feature of *when* is checked in Spec of CP against the +*wh* feature of C. If *when* or C can not check their +*wh* feature, the derivation crashes, as in 2.35b. Minimalism distinguishes between strong features, which must be checked in overt syntax, and weak features, which due to another principle (i.e. 'Procrastinate') must be checked in covert syntax.[12]

While research in an MP framework still recognizes some role for parameters, it attempts to reduce the sources of parametric variation, and it works for explanations that provide only a negligible role for learning in the domain of syntax, with a far greater role for learning in the domain of the lexicon. MP has not fully discarded its GB predecessor, however, and the well-formedness of constructions is still seen to be a function of language-independent universal principles, together with parameters revealed in the acquisition of the lexicon, such as whether a language has strong features or weak features.

2.6 Optimality Theory and pro-drop

This section introduces Optimality Theory by first discussing the basic tenets of this theory and then illustrating the theory through a discussion of OT research to date pertaining to the three aspects of pro-drop that are the focus of this dissertation (null subjects, inversion, and *that-trace*). Given the extraordinary number of papers written on pro-drop from a mainstream generativist perspective, this chapter's review of that literature has been necessarily selective. There are far fewer OT treatments related to pro-drop, and the review here is a more complete representation of work done from that

perspective; however, three important papers are not included in this section: Baković (1997) will receive only brief mention in this section, but will also be discussed in a response to Newmeyer (2000) in Section 2.7; and two positions regarding the acquisition of null subjects (Park 2000 and LaFond, Hayes, and Bhatt 2000) will be reviewed in Chapter 3, which deals with second language acquisition.

Like MP and its predecessors, Optimality Theory (Prince & Smolensky 1993, Grimshaw 1997) is a theory of generative linguistics which assumes that humans are endowed with an innate capacity for language and a language-learning mechanism that exploits this capacity; however, unlike MP and its predecessors, OT proposes that what is universal in human languages is a shared but finite set of grammaticality constraints. More importantly, these constraints are all violable and potentially in conflict with one another, resulting in a hierarchy between the constraints that varies from language to language. In OT, therefore, language variation is not the result of different parametric settings, but the result of differences in the language-specific rankings of the same set of universal constraints. This means that the available candidates for the realization of a structure directly compete, and the candidate that wins is the one that is most harmonic, or optimal, with respect to the hierarchy of conflicting requirements in the grammar. Since grammars contain universal constraints that conflict with each other, the grammars' job is to manage this conflict. The mechanism involved in this management is graphically depicted in Figure 2.2:

Figure 2.2. Operation of an OT Grammar

| Input | → | GEN | → | Candidate set | → | EVAL | → | Optimal candidate |
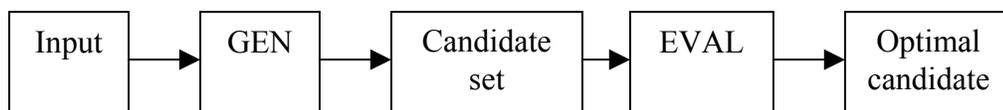
As hearers receive input, a candidate generator (GEN) produces the set of all possible candidate structural descriptions of that input.[13] A candidate evaluator (EVAL) takes this set of candidates and assesses them in light of the ranking of universal constraints particular to the language that is being learned. EVAL determines which candidates involve lesser or greater violations, given that particular language-specific ranking. The structural description that involves the least serious combination of violations is the optimal candidate. In such a system, there is always at least one candidate that is better than others, the candidate that has the fewest or least critical constraint violations.

Violations of constraints in OT are normally illustrated through the use of a 'tableau' (Tableau 2.1). Definitions of the constraints in Tableau 2.1 are unimportant for now; the tableau is used for illustrative purposes only. Constraints are ranked across the top, with the higher-ranked constraints to the left and the lower-ranked constraints to the right. Solid lines between two constraints indicate crucial rankings between those constraints; the use of a dashed line indicates that the constraints on either side are not ranked with respect to each other.

Tableau 2.1 Sample tableau adapted from Samek-Lodovici (1996:43)

<sing (x), x=*John*, -, T=pres.perf>	PARSE	DROPT	SUBJ	FULLINT
a. K *he has sung*				
b. *has sung he*			*!	
c. __ *has sung*	*!			

Candidates are given in the left hand column, with the optimal candidate (that with the fewest and/or least serious violations) marked by use of the 'K' symbol. All violations

are marked with an asterisk (*), and an exclamation point (!) after an asterisk marks that violation as 'fatal', meaning that the constraint that is violated is one which eliminates the candidate from further consideration as an optimal candidate, due solely to the fact that another candidate exists which does not violate constraints as highly ranked as the one marked with (!). Areas that are shaded mark constraints that are no longer relevant, since a higher ranked violation has already eliminated the candidate. Finally, the top left column of the tableau may contain the input to the system, as it does here.

Thus, in OT acquiring a language does not mean learning the well-formedness constraints of these structural descriptions (UG provides that); rather, language acquisition involves learning the language-particular ranking of these constraints, from strongest to weakest. When these rankings vary between languages, as in Figure 2.3, acquiring the second language will involve some type of reranking.

Figure 2.3 Differences in the ranking of a hierarchy

Hearer's Hierarchy   Constraint 1 » Constraint 2 » Constraint 3
Speaker's Hierarchy Constraint 2 » Constraint 1 » Constraint 3

Hutton (1996), in a discussion of sound change, provides a useful summary of the types of alterations to a constraint hierarchy that might occur 2.4:

2.4 Alterations to a constraint hierarchy (adapted from Hutton 1996:4)

a. Promotion of constraints
b. Demotion of constraints
c. Creation of new connections between constraints (A,B→ A»B)
d. Dissolution of connections between constraints (A»B→ A,B)
e. Alteration of dominance relationship between two constraints (A»B→ B»A)

As a theory of constraint interactions, OT differs significantly from the earlier parameter-setting framework. As stated above, in OT there are no inviolable principles, other than those that create the structure of the language-learning mechanism itself; there are also no binary parameters or conditions that depend upon the strength or weakness of features. Since violability of constraints is an essential feature of an OT grammar, constraints may be stated in their simplest version, without the use of 'hedges' to cover exceptions to what might be otherwise thought to be a universal principle.

One of the earliest OT syntax treatments, Grimshaw and Samek-Lodovici (1995), provided an analysis of null subjects and inversion (two pieces of the pro-drop 'parameter') that captured differences between Italian and English by looking at the interaction of syntactic and discoursal constraints governing topics and foci in these languages. Grimshaw and Samek-Lodovici demonstrated that referential pro-drop is restricted to topic constituents, and they also showed that the fact that subjects can appear post-verbally in Italian, but not English, also derives from the language-specific ordering of the same universal constraints for these two languages. Crucially, because of the approach that Grimshaw and Samek-Lodovici take, they are able to make robust claims regarding not only regarding which languages will or will not permit null subjects and free inversion, they are also able to make language-internal, construction-specific predictions based on the grammatical effects of the discourse status of arguments.

Grimshaw and Samek-Lodovici specify both some of the universal constraints governing subjects and discourse, and the rankings of these constraints in Italian and English (Tables 2.2 and 2.3):

Table 2.2 Constraints in Grimshaw & Samek-Lodovici (1995)

Constraints	Definitions
SUBJECT (SUBJ)	The highest A-specifier in an extended projection must be filled.
FULLINTEPRETATION (FULLINT)	Parse lexical conceptual structure.
DROPTOPIC (DROPT)	Leave arguments coreferent with the topic structurally unrealized.
ALIGNFOCUS (ALIGNF)	Align the left edge of focus constituents with the right edge of a maximal projection.
PARSE	Parse input constituents.

Table 2.3 Constraint rankings for English and Italian

Language	Ranking
English	PARSE » SUBJECT » FULLINTERPRETATION » DROPTOPIC » ALIGNFOCUS
Italian	FULLINTERPRETATION » DROPTOPIC » PARSE » ALIGNFOCUS » SUBJECT

Augmenting earlier research on null and inverted subjects, they do what previous accounts have not — they show the interface between syntactic and discoursal requirements. In so doing, they go beyond the claim that languages may typologically be divided into pro-drop and non-pro-drop languages to make the additional claim that it is possible to specify the environments where inversion and null-subjects will occur. Grimshaw and Samek-Lodovici begin with the observation that in Italian foci are right-adjoined (2.36):

(2.36.) Q: *Chi ha gridato?*   'Who screamed?'
        A: a. *Gianni.*   'Gianni screamed.'
              b. *Ha gridato Gianni.*
              c. **Gianni ha gridato*
              d. **Ha gridato, Gianni.*

While it is true that English and Italian differ in that Italian allows postverbal subjects and English does not, it is not true that subjects in Italian 'freely' undergo inversion.

Grimshaw and Samek-Lodovici maintain that subjects can only appear (and *must* appear) at the right edge of the VP when they are contrastively focused. In Italian, they claim, the ALIGNFOCUS constraint dominates a SUBJECT constraint. Therefore, when a subject is contrastively focused, it must move from its canonical subject position to the right edge of the VP. In English, SUBJECT outranks ALIGNFOCUS. Therefore, the subject does not leave its canonical subject position, even when contrastively focused. As Grimshaw and Samek-Lodovici note, when constraints related to discourse or information structure are allowed to interact with syntax, it becomes possible to see a particular instantiation of a sentence as an 'optimal form of a distinct input, related to the pragmatics of correction-contexts' (1995:597).

Grimshaw and Samek-Lodovici view OT as an improvement over parameter-setting frameworks that simultaneously over- and under- generate topic-dropping due to a formulation of the parameter that does not allow it to interact with different levels of language. By using interactions between violable constraints, Grimshaw and Samek-Lodovici's approach allows them to delimit the contexts where null subjects and focus-adjoining occur. Since the languages in question vary in terms of the dominance configuration between competing syntactic and discoursal constraints, such an approach is needed to account for the data.

Samek-Lodovici (1996) extended the earlier work with Grimshaw, continuing with constraints involving a competition between discourse and syntax. Samek-Lodovici's main concern was subjects and the constraints related to their surfacing in the

36

grammar. For Samek-Lodovici, null subjects appear due to unparsed overt subjects; they are not null in the input itself. Therefore, a null subject results only when a discoursal constraint requiring the dropping of topics dominates constraints requiring subjects to be structurally realized. This is seldom the case in English (which is a primary way in which English differs from pro-drop languages such as Italian or Spanish), but Samek-Lodovici concedes that DROPTOPIC could be a gradient constraint, and suggests that another version of DROPTOPIC might be DROPTOPIC$_{rel}$, a constraint requiring that arguments with topic antecedents be realized only minimally. Under the assumption that constraints from differing levels of grammar (phonological stress, discourse, syntax) can and do compete with one another, Samek-Lodovici maintains that stressed and unstressed pronominals will incur a differing number of violations of the DROPTOPIC$_{rel}$ constraint when the input is a topic-referring subject (Tableau 2.2):

Tableau 2.2 English topic-referring subjects (from Samek-Lodovici 1996:56)

$<sing$ (x), x=$John_{top}$, T=pres.perf>	PARSE	DROPT$_{rel}$	SUBJ	FULLINT
a. K *he has sung*		*		
b. *HE has sung*		*!*		
c. __ *has sung*	*!		*	

In a pro-drop language such as Italian, the constraints would be ranked DROPTOPIC$_{rel}$ » PARSE » SUBJECT, and this order would still favor null subjects over structurally realized subjects.

In another account, Speas (1997) returned to the issue of null subjects using OT (recall that she had earlier (1994) provided an MP account of null subjects). Speas claims that all languages have unpronounced pronouns whose reference must be understood within a context. Whereas in previous accounts, the occurrence of these pronouns was

subject to licensing conditions, and the interpretation of them subject to identification conditions, in Speas (1997) the crosslinguistic facts are derived by the ranking of three constraints, CONTROL, FREEPRONOUN, and MAX(PRO). Definitions of these constraints are given in Table 2.4. The first two definitions are unhedged versions of principles in Principles and Parameters Theory; the third constraint is added by Speas.

Table 2.4 Constraints used in Speas (1997)

Constraint	Definition
CONTROL:	A null pronoun must be controlled in its control domain.
FREEPRONOUN:	A pronoun must be free in its governing category.
MAX(PRO):	If pro occurs in the input, its output correspondent is pro.

Speas argues that in languages such as Thai or Korean, which allow both subject and object pro, FREE PRONOUN and MAX(PRO) are higher ranked constraints than CONTROL. In languages such as Spanish or Mandarin, which have subject pro but not object pro, CONTROL and FREEPRONOUN are also higher ranked than MAX(PRO). But to explain differences between English and Spanish, Speas (1997) uses additional constraints (gleaned from Samek-Lodovici 1996), NO PHI-FEATURES ('avoid agreement features') and AGR(X) ('a tensed verb should host spec-head agreement between an agreement feature x and a nominal constraint').

Speas holds that the constraint rankings for English and Spanish are the same in regards to CONTROL, FREEPRONOUN, and MAX(PRO). The difference between them comes from the ranking of NO PHI-FEATURES and AGR(X), where English ranks NO PHI-FEATURES higher than AGR(X) and Spanish does the reverse.

It should also be noted that Speas (1997) provided an early defense of the use of OT for syntax by critiquing the claim that syntax is made up of inviolable principles. Her

strong argument was that every principle in Principles and Parameters Theory, while claiming to be universal, actually involves hedging. She illustrates this graphically by showing the principle, its essence, and its hedge (Table 2.5).

Table 2.5 'Inviolable principles' and their hedges (adapted from Speas 1997:184)

Principle	Essence	Hedge
Extended Projection Principle	All clauses must have a subject...	...except for languages which lack overt expletives.
Case Filter	An NP must have Case...	...unless it is null.
Principle A	An anaphor must be bound in its governing category...	...unless it is one of a special class of anaphors that need not be bound.
Principle B	A pronoun must be free in its governing category...	...unless it occurs in an idiom like *lose her temper*.
Principle C	A name must be free...	...unless its an epithet.
Empty Category Principle	A trace must be properly governed...	...where 'proper government' means government by a lexical head or by a close enough antecedent
Theta Criterion	All thematic roles must be assigned to an argument position, and all argument positions must receive a thematic role...	...except that the agent of a passive may be absorbed by the verb, and the thematic roles of nouns need not be syntactically realized.

In another account, Grimshaw (1997) provides the beginnings of an explanation of *that-trace* effects in English. To do so, Grimshaw uses two government constraints (Table 2.6):

Table 2.6 *That-trace* constraints in Grimshaw (1997)

Constraint	Definition
T-GOV	Trace is governed.
T-LEX-GOV	Trace is lexically governed.

T-LEX-GOV is violated whenever the trace is not governed by a lexical head. T-GOV is violated only if a trace is not governed by any head. Grimshaw notes her indebtedness to Déprez (1994) not only for T-LEX-GOV constraint, but also for the insight that English *that-trace* sentences are ungrammatical due to an alternative, the *that*-less sentence. Grimshaw's remarks that the presence of 'that' does not prohibit the extraction of an adjunct or complement, but extraction of a subject is only possible when no 'that' is present.[14] Therefore, there is no problem with sentences such as those in 2.37 (Grimshaw's example 46):

(2.37.) a. *Who do you think (that) they will see t?*

  b. *When do you think (that) they will see them t?*

Extraction of the object in 2.37a involves no violation of either constraint, whether the 'that' is present or not. Extraction of the adjunct in 2.37b involves no government at all, resulting in both constraints being violated, whether the 'that' is present or not. The result is that both options (with or without the 'that') incur the same number of constraint violations for the extraction of objects (here both incur no violations) and for the extraction of adjuncts (here both incur two violations). The result is true optionality.

The situation changes when there is extraction of a subject. Consider the sentences in 2.38a and 2.38b (Grimshaw's example 47):

(2.38.) a. *Who_i do you think t_i will see them?*

b. **Who_i do you think that t_i will see them?*

This time the two candidates differ in the constraints they violate. *That* may govern the

trace, but since the complementizer position is not a lexical category, C does not

'lexically' govern. Therefore the candidate that includes *that* in 2.40b violates the T-

LEX-GOV constraint, while 2.40a incurs no violations. See Tableau 2.3:

Tableau 2.3. Extraction of a subject (adapted from Grimshaw 1997:416)

	T-GOV	T-LEX-GOV
a. K $\quad$ V [$_{IP}$ t_i I [$_{VP}$ t_i V ...]]		
b. $\quad$ V[$_{CP}$ *that* [$_{IP}$ t_i I [$_{VP}$ t_i V ...]]]		*!

Grimshaw's analysis is also used by Baković (1997) who does not essentially

disagree with Grimshaw (1997), but points out that in examples such as those given

above, the relative ranking of T-GOV and T-LEX-GOV plays no role; Since T-GOV is

satisfied by both lexical and nonlexical government, it is not a useful constraint for

distinguishing forms.

In place of T-GOV, Baković substitutes the constraint FAITH[SUB], stating that the

output value of [SUB] (for 'subordination') is the same as the input value. To explain

what is meant by this, Baković assumes that the distinction between an embedded CP and

an embedded IP is their specification for subordination (i.e. CPs are [+SUB] and IPs are

[-SUB]). Baković further assumes that the specification of [SUB] is provided in the input,

and that it is the constraint FAITH[SUB] that regulates differences between input and

output in terms of the [SUB] feature. If this is the case, then the violations incurred by

sentences in 2.39 would vary based on the input, illustrated in Tableau 2.4.

(2.39.) a. [*Which coat*]ᵢ *do you know* [CP *that tᵢ doesn't fit?*]

   b. [*Which coat*]ᵢ *do you know* [IP *tᵢ doesn't fit?*]

Presumably, Baković's constraints would yield different results if ranked

differently in another language (e.g. Spanish), but his analysis does not show this.

Tableau 2.4. Constraints on *that-trace* (Baković 1997)

Input: [+SUB]		T-Lex-Gov	Faith[Sub]
a	[*Which coat*]ᵢ *do you know* [CP *that tᵢ doesn't fit?*]	*!	
b. ☞	[*Which coat*]ᵢ *do you know* [IP *tᵢ doesn't fit?*]		*
Input: [-SUB]		T-Lex-Gov	Faith[Sub]
a.	[*Which coat*]ᵢ *do you know* [CP *that tᵢ doesn't fit?*]	*!	*
b. ☞	[*Which coat*]ᵢ *do you know* [IP *tᵢ doesn't fit?*]		

Finally, Costa (1997) analyzes a number of constraints that have been proposed in

relation to focus and canonical word order. Costa observes that some languages (e.g.

English) use intonation for their primary focusing strategy, while others (e.g. Portuguese)

use word order. The examples in 2.40 demonstrate (using capital letters to indicate

primary sentence stress) how intonation indicates focused constituents in emphatic

sentences in English (adapted from Dalbor 1997:68):

(2.40.)   a. *ALEX will never do that!*

   b. *Alex will NEVER do that!*

   c. *Alex will never do THAT!*

   d. *Oh, yes, Alex WILL do that!*

This is at variance with a language like Spanish, where word order changes to show sentence stress in emphatic utterances, because in these utterances the element with sentence stress is placed on the right edge of the sentence 2.41:

(2.41.) a. *¡Nunca lo hará ALEJANDRO!*   (from Dalbor 1997:68)

   b. *¡Alejandro no lo hará NUNCA!*

   c. *¡Alejandro nunca hará ESO!*

   d. *¡Ah, sí, Alejandro lo HARÁ!*

English syntactically requires an object NP, a direct complement of the verb, to appear adjacent to the verb, even when it is focused and would be more appropriately placed (from a discourse perspective) in the clause-final position. This differs also from Portuguese, where the object must move to a clause-final position when focused, ignoring syntactic requirements to be adjacent to the verb. This is exemplified in 2.42 (from Costa 1997):

(2.42.) a: *Who does Paul know well?*   a: *Quem é que o Paulo conhece bem?*
     b: *Paul knows MARY well.*   b: **O Paulo conhece a MARGARITA bem.*
     b': **Paul knows well MARY.*   b': *O Paulo conhece bem a MARGARITA.*

In the Portuguese version, 2.42b' is correct only when it is the object that is focused. Costa (1997b) shows that a shift of focus to the adverb *bem* would reverse the acceptability of 2.42b and 2.42b'. This could occur, for example, after a question such as, *Como é que o Paulo conhece a Margarita*? 'How does Paulo know Margarita?'.

   Costa (1997b) uses Portuguese, English, Dutch, and Icelandic to illustrate differences between languages where there is a correlation between word order and

discourse functions and languages where this correlation is missing. Costa observed that Portuguese subjects and Dutch and Icelandic objects are generated in a position where they are focused regardless of the need for Case licensing. English, in contrast, violates discourse-related constraints when an element needs Case. Costa's observation challenges the Principles and Parameters framework (which would say that Case is an inviolable principle) while it provides strong evidence for discoursal and syntactic interactions of the same elements being resolved differently in differing grammatical systems.

2.7 A response to Newmeyer (2000)

The previous section introduced OT as a promising approach for dealing with interactions between discourse and syntax, pointing out that, although Optimality Theory (OT) is a generative approach (since it assumes learners possess an innate language acquisition device and that individual grammars are acquired as this device encounters linguistic input), this theory distinguishes itself from other generative approaches by viewing the universal principles of language as violable, and potentially conflicting, constraints.

OT was chosen as the theory of grammar that informs the analyses given in this dissertation for two chief reasons. First, OT provides a precise way to address 'the problem of language acquisition'.[15] Current research (e.g. Tesar and Smolensky 2000, discussed in this section and in the next chapter) has directly focused on the question of 'learnability' and provided clear demonstrations that OT yields a learnable model of grammar. Second, the structure of OT permits constraints from various levels of

grammar (phonology, syntax, discourse, etc.) to vie for prominence in the determination of the grammaticality of an output.

Both of these avowed reasons for using OT are controversial. Mainstream generativists have challenged the claim that OT grammars are sufficiently restrictive to limit the class of possible languages. They contend that an OT grammar will ultimately fail to produce 'all and only' the grammatical choices in a given language, that it will produce 'rogue grammars', and that total hierarchies of 'unmotivated competition' result in grammars that are unlearnable. Furthermore, they have argued that using OT to address issues of syntax[16] is an enterprise that is beleaguered by inherent and irresolvable problems, at the heart of which is the OT view that a grammar consists of a complete hierarchy of competing constraints. For critics, this view leads to a failure to respect the autonomy of different grammatical components, particularly 'the autonomy of syntax'.

The position that different components of grammar must be analyzed independently from the rest of the grammatical system is at the center of a longstanding linguistic debate between formalists and functionalists.[17] In this debate, formalist theories (e.g. Government and Binding, Principles and Parameters, Head-Driven Phrase Structure Grammar, Generalized Phrase Structure Grammar, Lexical-Functional Grammar, Minimalism, and in part Relational Grammar) have differentiated relationships between grammatical elements and sought to keep analyses fully separate from any semantic, pragmatic, or discoursal properties of those elements. In contrast, functionalist theories (e.g. the Competition Model, Role and Reference Grammar, Cognitive Grammar, Systemic Grammar, Emergent Grammar, Functional Grammar, and Construction Grammar) have largely rejected the compartmentalization of grammatical

45

form, insisting that the function of assigning and conveying meaning involves linking semantic and pragmatic elements to other formal parts of the grammatical system (e.g. syntax). Both camps agree that formalists are ignoring certain facts in the interest of articulating a theory; formalists embrace this idealization while functionalists seek ways to overcome it.

At least one common goal for both camps is to provide an account of the linguistic knowledge of native speakers of a language. Creating such an account is a complex task and has led the formalists to, as Culicover (1997:11) has noted, either ignore 'certain facts that do not naturally fit' or 'decompose phenomena into components' under the theory that it will be easier to deal with the components of language as isolated parts than as an undifferentiated whole. That the components of grammar are legitimately analyzed in isolation from one another has been the view of many since Chomsky (1975) sought to demonstrate that meaning plays little or no role in syntactic analysis. Chomsky's infamous phrase, 'Colorless green ideas sleep furiously', illustrated the point. The fact that native speakers of English identify this sentence as grammatical, despite its semantic gibberish, is thought to demonstrate that syntax is fully separate from semantics.

Hale, Jeanne, and Platero (1977) also developed this idea of autonomy in their Modularity Thesis or Autonomous Systems view that held that generalizations regarding one component of grammatical structure are statable without reference to interpretation or use. In addition, Jackendoff (1983) articulated the same view, maintaining that the set of components for each level of the grammar (and here 'grammar' refers *only* to syntax,

semantics, morphology, or phonetics/phonology) have their own well-formedness conditions, primitives, or rules of combination.

It is easy to see that in a system where the autonomy of components is assumed, there is no place for the primitives or well-formedness conditions of one component of grammar to make any reference to aspects of representation of another component of grammar. Although correspondence rules may map representations at one level onto representations of another level, there is no sense in which a phonological or semantic (let alone a discoursal or pragmatic) rule might determine the syntactic acceptability or grammaticality of a structure.

As Newmeyer (1998) points out, the 'autonomy of grammar' argument itself decomposes into three independent hypotheses: first, that there is a system of primitive terms whose combination makes no reference to semantics, discourse, or system external factors (autonomy of syntax); second, that the language competence is not dependent on language use or performance (autonomy of knowledge of language); and third, that the cognitive system of grammar is distinct from other cognitive systems (autonomy of grammar as a cognitive system). Newmeyer (1998) assembles empirical evidence in favor of each of these hypotheses. To show that syntax is independent of discourse function, he looks at various syntactic principles (*wh*-movement, lexical government, SUBJ/AUX inversion, etc.) for which he claims there is no direct link between form and function. He shows, for example, that SUBJ/AUX inversion takes place in requests, offers, exclamations, and questions, such as those shown in 2.43a-d.

(2.43.) a. *Could you pass the peas, please?*

   b. *May I help you?*

   c. *Is that ever a beautiful sky!*

   d. *Could South Carolina have a winning season?*

According to Newmeyer (1998), the fact that constructions such as these have no uniform semantic properties poses a challenge to semantic or pragmatic accounts. He concludes, 'The grammatical properties of human language are best characterized in terms of autonomous formal systems' (1998:365).

Newmeyer is correct in arguing that OT syntax violates the principle of syntactic autonomy; but the last 40 years of syntactic theorizing within mainstream generativist theory have demonstrated that analyzing grammatical components in isolation does not always result in more elegant or adequate explanations of linguistic data. Recent research even outside of OT (e.g. Liceras 1988, 1989, Liceras and Díaz 1995, Pérez-Leroux and Glass 1997, 1999, Pérez-Leroux, et al. 1999), has begun to recognize that learner grammars are not comprehensible without an examination of semantic and discoursal interfaces with syntax. Within OT, claims of interactions between levels of grammar are even more pronounced. The analysis given in Chapter 6 of this dissertation, operates from the assumption that there are simply too many facts that must be ignored if interface grammars are disallowed from the onset.

Newmeyer believes that the use of OT for purposes of syntactic analysis is 'seductive' because, while it demands precision (in that ranking constraints demands greater formalization than is found in some other generativist theories), it is also more flexible (in that constraints may be not only violable but also invented or ad hoc).

Nevertheless, Newmeyer believes this enticement should be resisted because OT syntax suffers from three inherent flaws: (1) The restrictiveness problem; (2) The functionality problem; and (3) The problem of unmotivated competition. Since the SLA program described in this dissertation is based upon an OT syntax approach, each of these arguments must be addressed.

2.7.1 The restrictiveness problem

Newmeyer does not believe that OT grammars are able to exclude logically possible, but naturally nonoccurring, language types, regarding as baseless the claims in many OT papers that free constraint ranking is capable of explaining typological distributions of grammatical elements. He believes that such claims can only be made because OT researchers presuppose that the only relevant constraints are the ones being addressed in that specific paper. As evidence for his critique, he argues that Grimshaw and Samek-Lodovici's (1998) proposal that languages may have null topics as subjects, but not null contrastively focused subjects, falls apart if one adds a constraint he calls DROPFOCUS to the DROPTOPIC constraint used by Grimshaw and Samek-Lodovici. He defines this constraint as in 2.44 (Newmeyer 2000:11):

> (2.44.) DROPFOCUS: Leave arguments coreferent with the focus structurally unrealized. [Failed by overt constituents which are coreferential with the focus].

Newmeyer believes such a constraint must be valid because in some languages focus can be pragmatically retrievable. As an example, he presents the following, where an

exchange such as 2.45 in English could be communicated via a dropped focus constituent

in Portuguese 2.46 (Newmeyer 2000:12):

(2.45.) Q: *Who broke the plate?*
A: *You know who!*

(2.46.) Q: *Quem é que partiu o prato?*   'Who is it that broke the plate*?*'
A: *Partiu o prato?*   'Broke the plate*?*'

Newmeyer also suggests that given a particular question intonation, this type of 'focus-

dropping' could also be possible in Serbo-Croatian. Newmeyer concludes, 'Grimshaw

and Samek-Lodovici are incorrect in claiming both that they (null focused subjects) *don't*

occur and, given the architecture of OT, that they *can't* occur' (2000:14). He further

asserts the following:

> One is never on safe ground in claiming that one's particular OT analysis is
> 'restrictive' in any interesting sense of the term. The limited set of constraints
> exhibited in the tableau in a particular analysis might well seem to conspire to
> predict the impossibility of some typologically-dispreferred structure. But since
> every tableau (when fully expanded) contains every constraint, it seems
> inevitable that *some* constraint will always lurk in the background which, if
> ranked highly in some particular grammar, will lead to the generation of that
> dispreferred structure.
>
> (Newmeyer 2000:14)

If this critique stood, it would clearly be damaging to the OT program; however,

there are several sensible ways to respond to this criticism. First, one must surely

question whether the interpretation that Newmeyer wishes to place on the exchange in

2.48 is syntactically encoded. OT does not claim to account for all idiomatic, non-

productive interpretations that are highly context-dependent; neither does any other

grammatical theory to date. If Newmeyer wishes to create a unique set of discourse

conditions, with a particular type of question intonation, together with a great deal of

shared contextual knowledge between the participants, and then argue for a particular

50

grammatically encoded interpretation, he is claiming far more than any OT analysis to date has done.

Second, Newmeyer is surely correct that restrictiveness is a serious problem if constraints can be constructed willy-nilly in the manner of his DROPFOCUS example. Dropping focused constituents is not a general or motivated constraint for any language, including Portuguese and Serbo-Croatian. In contrast, the dropping of topic subjects is a well-attested phenomenon in a great number of adult languages, and even more universally attested in the language development of children. Newmeyer's critique rightfully demonstrates the negative impact of the ad hoc construction of constraints, but his critique does not pose a serious challenge to the feasibility of ranking well-motivated and broadly attested syntactic and discoursal constraints with each other.

Third, one must question whether Newmeyer's specific concerns about restrictiveness are as much of a problem for OT as they are for the Principles and Parameters framework that he assumes. This problem relates ultimately to the question of learnability, and though there is often an assumption that hypothesized grammars in language acquisition must be only those that are fully specified and admitted by UG, this is not a necessary conclusion. Tesar and Smolensky (2000), for example, point out that while such an assumption has the advantage of explaining why adult grammars invariably end up in the UG-allowed space, it has the potential disadvantage that it requires learners to be fully committed to a particular dimension of grammatical variation for which they have insufficient evidence. Such a position also does not deal as appropriately with adult L2 learner grammars, grammars that can, and often do, end up at variance with adult native speakers of the target language.

Finally, although in generativist circles it has become dogma that the job of linguistic theory is to restrict the space of grammars, this has popularly amounted to nothing more than delimiting a finite set of possible parameters with the associated assumption that the fewer possibilities, the better (i.e. more learnable) the theory. This popular notion is not informed by learnability theory. As Tesar and Smolensky (2000:2-3) have noted:

> ...limiting the set of possible grammars to a finite number serves only to improve the worst-case performance of the *least informed* learning method of all: exhaustive search, in which every possible hypothesis is examined...a grammatical theory with an infinite number of possible grammars might be well structured, permitting *informed* search that converges quickly to the correct grammar — even though uninformed, exhaustive search is infeasible.

Some adherents of PPT have not been adverse to invoking whatever new parameter seems to fit the needs of specific analyses, calling in question the 'finiteness' even of parameter theory, but as Tesar and Smolensky point out, even limiting a grammar to a finite number of parameters does very little to improve the learnability of a grammar, if the total possible number of grammars is quite high. Tesar and Smolensky deduce, '...a well-structured theory admitting an infinity of grammars could well be feasibly learnable, while a poorly structured theory admitting a finite, but very large, number of possible grammars might not' (2000:2).   Whether a grammar is ultimately learnable depends more on the learning mechanism than on the total number of candidates in a search space.

Learning research conducted in connection with principles and parameters (e.g. 'cue learning,' Dresher and Kaye 1990, or the 'triggering learning algorithm,' Gibson and Wexler 1994) has moved in one of two directions, both of which attempt to respect the autonomy of grammatical components: either they narrowly use the grammatical structure of a specific parametric system, or they use algorithms so general that they

apply to any (even nonlinguistic) parametric system. As Tesar and Smolenky (2000:4) note, such attempts come as a result of the fact that parametric systems have 'little structure for the learner to exploit beyond the existence of a finite space for learning.' In contrast, the OT approach advanced in this dissertation is a theory of crosslinguistic variation that uses a grammatically-informed learning algorithm that is demonstrably learnable.

Therefore, one may legitimately question whether the 'problem' of restrictiveness is a real issue or a pseudo-concern that arises from a particular set of generativist beliefs. It is true that OT allows a very large search space indeed. In fact, the number of potential grammars is equal to the number of possible total rankings (*N!*). But a process such as Constraint Demotion (Tesar and Smolensky 2000) drastically reduces the data complexity imposed by a large number of constraints, and this permits a useful type of restrictiveness not necessarily present in other generativist approaches (i.e. learners can efficiently converge upon target grammars through the use of the inherent structure provided by strict domination).[18] Principles and parameters, despite its more 'restricted' learning space, has yet to demonstrate that it provides an approach to language acquisition that is actually learnable.

2.7.2 The functionality problem

Newmeyer's second critique of OT is that some have attempted to use it to identify functional motivations for each of its universal constraints. Since Newmeyer rests firmly within the formalist camp, it is unsurprising that he calls this attempt 'fundamentally wrong-headed'. Nevertheless, as noted earlier, researchers operating

within an OT orientation do span both sides of the functionalist/formalist divide. This is a natural outcome in that OT provides no fully elucidated theory of the universal constraints, but rather takes the constraints of particular grammatical theories and provides an organizing principle for them. For example, the SLA approach in this dissertation proceeds from a generativist view, assuming that candidate structures (constructed by GEN and considered by EVAL) conform to the basic architecture of the X-bar theory developed by Jackendoff in the 1970s and incorporated into GB theory in the 1980s. It differs from other generativist approaches, however, in that it shares with functionalism a rejection of the autonomy of syntax and a full appreciation for the interfaces that exist between grammatical components.

Nevertheless, there are those papers that view OT from a stronger functionalist orientation (e.g. Haspelmath 1999, Aissen 1999) and attempt to bridge formal and functional linguistics by finding constraint-function pairings that root OT constraints in their functional utility. Newmeyer's critique here is not objectionable, and even he grants that functional motivations could play a role in language use and acquisition (with the assumption that questions of acquisition and use are outside the core of linguistic theorizing). Although considering why the constraints are the way they are may be an interesting diversion, these types of justifications are irrelevant if the theory assumes the innateness of constraints (as even Haspelmath 1999 admits). Functional motivations play little role in the work of the research program covered in this dissertation, and it is reasonable to concede the problem that Newmeyer raises regarding functionality, with the caveat that it applies only to some who work within OT syntax, and is not a failing of the theory itself.

2.7.3 Unmotivated competition

Newmeyer's final criticism is his strongest attack on OT. Newmeyer rightly
asserts that OT grammars fail to respect the autonomy of different grammatical
components, a failure, he maintains, that obscures important generalizations about
grammatical patterning. He labels this the problem of 'unmotivated competition' because
it places grammatical components from differing levels in direct competition with one
another. Newmeyer rightly observes that this competition is at the heart of OT, and is
used to show that variances in hierarchical rankings account for differences between
languages. However, he believes that this core notion leads to 'ad hoc competition sets',
'uninsightful analyses', and, even worse in his estimation, the importation of discourse
constructs into syntax to handle discourse-dependent optionality, violating the received
doctrine of syntactic autonomy. Newmeyer bases his argument on analyses of two types
of optionality, discourse-independent, illustrated in 2.47 and discourse-dependent,
illustrated in 2.48 (Newmeyer 2000:46).

> (2.47.)   a. *I believe that it will rain tomorrow.*
>   b. *I believe it will rain tomorrow.*

> (2.48.)   a. *You will never get me to eat fruit-flavored tofu.*
>   b. *Fruit flavored tofu you will never get me to eat.*

An OT grammar must be able to show how each of these types of optionality can arise.
How it can accomplish this is not obvious, however, since OT grammars are
deterministic, mapping each input into a single output that possesses the greatest
harmony in relation to the ranking of constraints. If 2.47a and 2.47b are different in
grammatical terms, and they are both generated from a single input, then the differences

leading to the inclusion or the exclusion of *that* must be due to constraints in the hierarchy.[19] The choice of one or the other of these candidates appears to require that one is more optimal than the other, but every intuition of native speakers suggests they are in free variation.

OT researchers have addressed the issue of optionality in a variety of ways. Kager (1999:405) mentions that one terminological way of dealing with the issue is to redefine an output of the grammar as a set of forms rather than a single form, but concedes, 'this does not solve the problem of how to generate sets of output forms'. Another approach, advocated by Legendre (2000) is based on the claim of Kroch (1989) that individuals have multiple grammars for their different stylistic registers. Legendre (2000) shows that in French *wh*-constructions, level of formality determines the relative ranking of some constraints. Newmeyer opposes such an approach on the grounds that generative syntax has traditionally treated data from various registers of a grammar together, as with the *wanna*-contraction or stranded prepositions. This is, of course, a very weak argument — that researchers should reject this approach because it is has traditionally been done another way.

True optionality can exist in OT if two constraints are crucially unranked in relation to each other. For example, assume that language L has constraints w, x, y, and z. If the learner has received input that would require the demotion of say, x below the other three constraints, and then further received input that would demote y below the remaining two constraints, and then further received input that would demote y below x, the constraint hierarchy would look as follows: $w, z \gg x \gg y$. This hierarchy leaves w and z unranked in relation to each other, leaving more than one possible choice as optimal in

this grammar. This type of free ranking permits an undetermination of the grammar. Kager (1999:407) notes that accounting for free variation may lead OT further in the direction of connectionism, where a numerical index indicating relative strengths in relation to other constraints may provide a probabilistic view of constraint interaction. This approach is similar to that of Anttila (1997), who also believes that variation materializes when the grammar underdetermines the output. As some researchers have argued, and I will later schematically illustrate, partial ordering of constraints is a possibility. For Anttila, *preferences* occur if these partial orderings are too weak to select a single winner, but still strong enough to leave a type of statistical fingerprint on the output. Preferred forms are then the result of one candidate beating another in a greater number of tableau, but still leaving the possibility for realization of the variant form.

Newmeyer dismisses true optionality as theoretically unlikely because 'it is so easily sabotaged by the existence of some low-ranked constraint that distinguishes the two variants' (2000:49), but a mere dismissal does not amount to evidence against it; furthermore, while it is true that in a totally ranked hierarchy (with no boundaries between components of grammars) such true optionality would be rare, the possibility for this type of optionality becomes far greater if the language learner does not have to consult the entire hierarchy for each piece of data, but is able to focus on the portion of the hierarchy relevant to a given structure. Exactly how this may be done within the learning algorithm is not fully clear, but the possibility of such a solution surely militates against viewing OT syntax as 'inherently flawed'.

Numerous other possible solutions to optionality are debated among OT researchers, including ordered global constraint ties (2.49), ordered local constraint ties

(2.50), conjunctive local constraint ties (2.51), local disjunctive ties (2.52), and global

disjunctive ties (2.53), here illustrated schematically (as adapted from Newmeyer 2000):
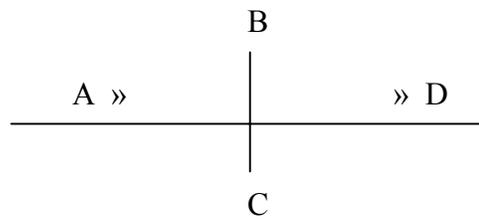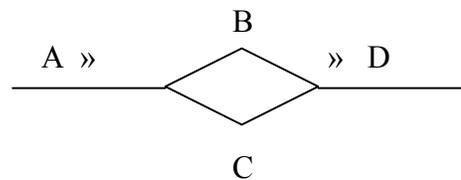
(2.49.) Ordered global constraint ties

```
                »      C      »      D
          B
    A   »
                »      B      »      D
          C
```
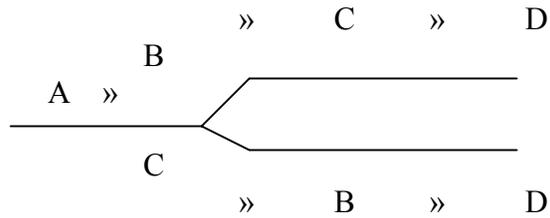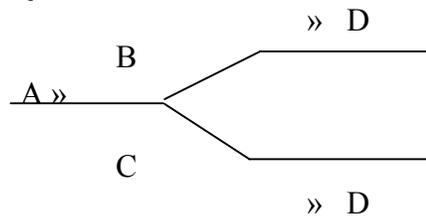
(2.50.) Ordered local constraint ties

```
                       »
          B                 C
    A   »                        D
          C                 B
                       »
```
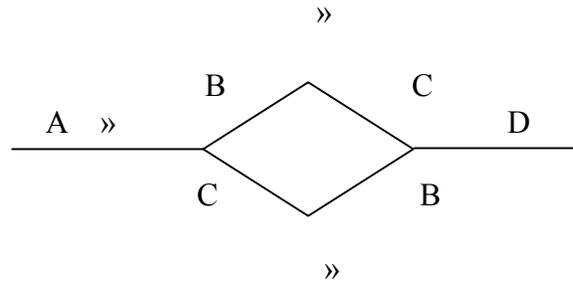
(2.51.) Conjunctive local constraint ties

```
                 B
    A »                  » D
                 C
```

(2.52.) Local disjunctive ties:

```
                 B
    A »                » D
                 C
```

(53.) Global disjunctive ties

```
                                      »  D
              B
     A »
              C
                                      »  D
```

The details of each of these theories are not critical here. The basic idea behind them is

that crucial non-ranking of certain constraints in a grammar leaves room for some

optionality. Newmeyer does nothing to refute these proposals himself; he relies on the

debate between OT researchers on the advantages or disadvantages of certain approaches

to suffice. His own position is simply that lower-ranked constraints must somehow

destroy the possibility for an optimal choice, and that these types of approaches are

difficult to reconcile with the learnability algorithms such as that of Tesar and Smolensky

(2000). Once again, this type of argumentation does not approach any threshold of

evidence for the invalidity of an OT approach, and it fails to understands how constraint

demotion actually works. Constraint demotion does not function, as Tesar and

Smolensky (2000:48) state, within 'the confines of the space of totally ranked

hierarchies'; rather, the algorithm operates within the larger space of stratified

hierarchies.

'Unmotivated competition' between constraints is Newmeyer's concern wherever

the autonomy of syntax appears to be violated by OT. For optionality that does not

involve differences in the information structure of the utterances, Baković's (1997)

analysis of optional complementizer deletion may serve as an example. Baković

(1997:1) cites examples such as 2.54 and 2.55:

(2.54.) a. *The coat [**that** he always wears t] doesn't fit him.*
      b. *The coat [he always wears t] doesn't fit him.*
(2.55.) a. *I think [**that** the coat doesn't fit him].*
      b. *I think [the coat doesn't fit him].*

Sentences 2.54 and 2.55 show that complementizers in English are optionally present in some contexts.[20] Baković argues that if the only constraints in a hierarchy were markedness constraints, such optionality would not be possible — the constraints would always compete, with the least-marked form invariably winning. However, markedness constraints may interact with faithfulness constraints, and under such conditions, optionality is a natural result. If we assume that specification for a subordination feature [SUB] is a part of the input that is received, then a ranking of faithfulness (ensuring that input and output values remain constant) over other markedness constraints would result in the following hierarchy, where optionality is the result of variance in the input, not a variable ranking of constraints within the grammar (Tableau 2.5):

Tableau 2.5 Complementizer optionality

Input: [+SUB]	FAITH[SUB]	MARKEDNESS CONSTRAINTS
a. ☞ *I think that the coat does not fit him*		*
b. *I think the coat does not fit him*	*!	*

Input: [-SUB]	FAITH[SUB]	MARKEDNESS CONSTRAINTS
a. *I think that the coat does not fit him*	*!	*
b. ☞ *I think the coat does not fit him*		*

Newmeyer's critique of this type of approach is stated but not argued. First, he objects to the expansion of the input to include functional contrasts. Second, he opposes the interaction between faithfulness and markedness constraints on the basis that it involves larger candidate sets and creates 'vacuous ambiguity' in well-formed sentences.

Neither of these objections strike the death-blow to OT that Newmeyer supposes. First, there is no reason to assume, a priori, that functional information cannot be part of a syntactic input. Indeed, a number of analyses within the mainline generativist approaches require a refined sensitivity to functional categories (Liceras 1988, 1989, Liceras and Díaz 1995, Pérez-Leroux and Glass 1997, 1999, Pérez-Leroux, et al. 1999), and if this sensitivity is not developed on the basis of input, it is unclear where it comes from. Second, the increase that faithfulness constraints create for the complexity of the candidate sets is only of concern if the exhaustive search method is used, but as Tesar and Smolensky (2000) have shown, the learning algorithm is significantly more sophisticated than that. It may be that Baković's approach to handling discourse-independent optionality requires further development, but Newmeyer has not yet demonstrated this to be the case.

Newmeyer next considers 'discourse-dependent' or 'discourse-sensitive' optionality, optionality that occurs when the same input (which in OT syntax has been understood as lexical items together with their argument structure, tense, and aspect specifications)[21] results in variable outputs on the basis of information structure. Newmeyer's critique here is not unlike those already addressed. His primary concerns are that discoursal features are incorporated into the input, in the same way that information about functional categories was incorporated into the input in Baković (1997), and that the inclusion of such constraints expands the number of candidate sets beyond manageability.

For example, Newmeyer opposes Legendre's analysis of French Stylistic Inversion (from Legendre 1999)[22] on the grounds that allowing discourse or information structure constructs to compete with syntax poses a serious violation to the autonomy of

syntax. He argues that the principles characterizing well-formedness of sentences structurally must be different from those characterizing their appropriate use in discourse, using examples that place word stress on differing constituents, depending on the context, and arguing thereby that focus cannot be marked as a feature on a single constituent. He shows, for example that topics can be discontinuous (2.56), non-constituents (2.57), or not even need a syntactic representative (2.58):

> (2.56.) Q: *Are Mary and Tom friends?*
>   A: *No, Mary HATES Tom.*
>
> (2.57.) Q: *WHO did the man see?*
>   A: *The man saw THE WOMAN.*
>
> (2.58.) Q: *What's the weather?*
>   A: *It's RAINING.*

From this he concludes that topic is not a property that can be marked as a feature on syntactic constituents. Furthermore, Newmeyer argues that the relationship between syntax and information structure is indirect, with formal marking showing no need to correspond directly to information status. To illustrate, Newmeyer refers to Gundel's (1988: 214a) example of a formally definite NP that does not represent information known to the hearer (2.59):

> (2.59.) *I couldn't sleep last night because the neighbor's dog kept me awake.*

Syntax and information structure usually have more than a one-to-one mapping. For example, Newmeyer cites Prince (1998) to show that processes such as left-dislocation or topicalization have multiple functions.[23] Semantic structure can also diverge from both syntax and discourse, so in 2.60 two structurally different sentences

have the same information structure, although the same lexical item is focused in both

sentences (from Vallduví 1992:113):

        (2.60.) a. *The boss hates BROCCOLI.*

            b. *It is BROCCOLI that the boss hates.*

Given all of these considerations, Newmeyer draws the following conclusion (2000:60):

> The built-in need of OT to regard each grammatical sentence as a winner in its
> own competition set has led to analyses that import aspects of information
> structure into the syntactic derivation. But the complexity and indirectness of the
> interface between syntactic structure and information structure shows that such a
> move is ill-advised.

On the face of it, Newmeyer's critique appears very damaging to OT syntax and

to the course pursued in this dissertation. The OT analyses used in this dissertation are

guilty of precisely the critiques Newmeyer raises by grammatically encoding discoursal

constraints and viewing them as universal elements of the OT grammar, and also by

violating the autonomy of syntax by arguing that discourse and syntax interact as the

second language learner moves towards acquisition of the target language. This is done

despite the critique of Newmeyer for the following reasons. First, although it may be

granted that discourse function is not invariably tied to a single constituent; for any given

linear ordering of lexical items, there are clearly numerous focus possibilities, dependent

on the discoursal requirements of the context. When a lexical item such as 'broccoli' in

2.62 is able to receive focus in more than one position in the sentence, this does not

demonstrate that there is no interaction between components of grammar; rather, it shows

that the significant interaction in this case is between information structure and stress

assigned via the phonological constraints of that language. After all, not every significant

interaction is based on syntax. Rather than arguing for the autonomy of syntax by such examples, Newmeyer is further demonstrating the need to examine interfaces between grammatical components.

Second, Newmeyer's claim that interactions between discourse and syntax are complex and, at least apparently, indirect can also be granted. It is precisely this complexity, together with the artificial requirement to separate syntax from other components of grammar, that underlies the failure of the mainline generativists to be able to account for various grammatical phenomena, of which pro-drop is a prime example. The claim that allowing discourse to interact with syntax in the grammar makes the grammar too complicated, or the constraints too numerous, is countered by the charge that refusing to allow discourse to interact with syntax yields only an overly simplified account that is non-predictive and must fall back on optional rules that ultimately yield only description, not explanation. The presence of complex principles of interpretation relating syntactic structure and information structure is unsurprising, nor should the complexity of these principles dissuade researchers from making progress where possible.

Third, the observation that syntax and information structure usually have more than a one-to-one mapping does not argue against an OT approach. OT presents at least the possibility of analyzing, in a way that is not possible within some other frameworks, the host of constraint interactions that may demonstrate why syntax and information structure do not have one-to-one mapping in certain cases. At best, mainline generativist accounts yield no better explanations in these situations; at worst, they do not possess even the means for addressing them.

Finally, while Newmeyer's defense of the autonomy of syntax permits him to easily demonstrate that OT violates this generativist principle, his defense does not meet necessary criteria of explanation. The burden of proof for the principle of autonomy rests with those making the claim of autonomy. While Newmeyer alleges that this burden has been met, this assertion ignores data that do not fit the claim (such as that offered in the next section) and inadequately meets certain criteria of explanation. Contrary to Newmeyer's claim, OT interactions between constraints can be precisely formulated and they make specific predictions, demonstrating a linkage between cause and effect. Furthermore, they result in measurable typological consequences.

Interestingly enough, Newmeyer's critique wistfully longs for halcyon days when sentences displaying differences in free variation would be related via optional application of transformational rules. Even after the generative program in the 1970s led to a single movement rule, optionality could be handled by different applications of *Move-α* to a single D-structure, and interpretive principles applying at LF would leave the syntactic component untouched. Newmeyer views Minimalism, where all movement is obligatory, as a serious and unfortunate shift away from optionality. Apparent optionality in MP must now arise from two derivations from different candidate sets, a move which has its own unwelcome results. Therefore, according to him, both minimalism and OT share the problem of how to treat non-unique outputs. But returning to the optional application of rules is a dead-end street. Optional rules are neither predictive nor explanatory, and on this OT and Minimalism agree.

2.7.4 The syntax/discourse interface

The autonomy of syntax is an empirical issue; the fact that syntax often appears to make well-formedness judgments apart from other considerations does not establish autonomy. To be truly autonomous, it must *always* do so, and the failure of syntax to do this is can be demonstrated. Consider, for example, previous work in pro-drop, the focus of the research program discussed in this dissertation. If, as Newmeyer would have it, answers related to a syntax/discourse interface are not admissible as an explanation of the pro-drop phenomenon, some purely syntactic solution must be posited. Of course, this has been done. One standard proposal is that pro-drop and non-pro-drop languages differ in that non-pro-drop languages such as English have weak agreement features and do not permit the licensing of pro (Figure 2.5), but pro-drop languages such as Spanish have strong agreement features and do permit the licensing of pro at Spec of INFL (Figure 2.6):

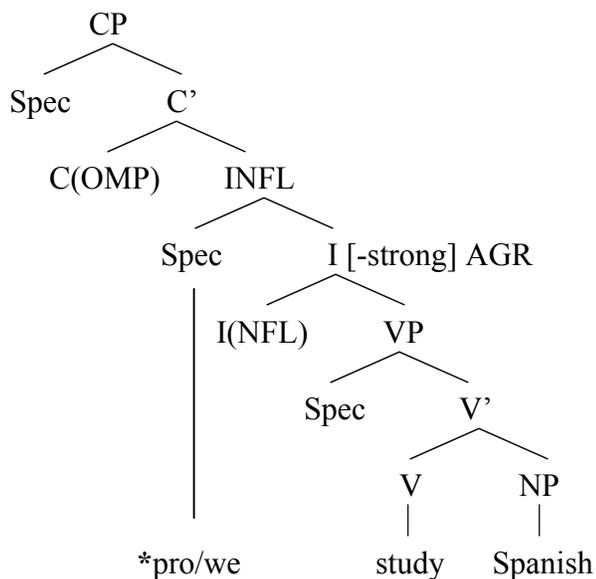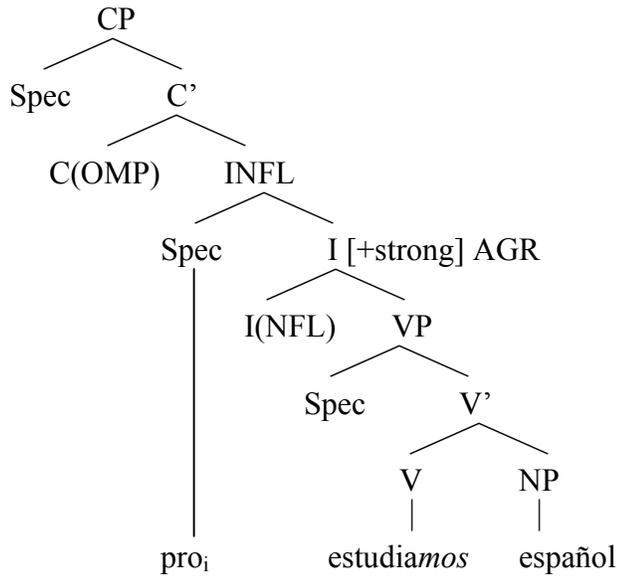Figure 2.5 Non-pro-drop languages: Weak agreement features

```
                CP
              /    \
          Spec      C'
                  /    \
             C(OMP)    INFL
                      /    \
                  Spec      I [-strong] AGR
                   |       /    \
                  I(NFL)  VP
                        /    \
                     Spec     V'
                            /    \
                           V      NP
                           |      |
        *pro/we          study  Spanish
```

Figure 2.6 Pro-drop languages: Strong agreement features

```
            CP
          /    \
       Spec    C'
              /    \
          C(OMP)   INFL
                  /      \
              Spec      I [+strong] AGR
               |        /      \
               |    I(NFL)     VP
               |             /     \
               |          Spec     V'
               |                  /    \
               |                 V      NP
               |                 |      |
             proᵢ           estudiamos  español
```

(From Liceras & Díaz 1999:10-11)

The problem with this type of analysis is two-fold: first, although this respects the autonomy of syntax, it is less descriptive and less explanatory than an account admitting discoursal factors. The explanation illustrated in Figures 2.5 and 2.6 does not account for other languages, such as Chinese, that have weak agreement and still have null subjects. Therefore, a disjoint solution becomes posited. In Chinese, so the suggestion goes (Huang 1984), pro is licensed by a discourse-bound operator in Spec of CP, and identified by null topics. So we now have three types of languages, those that license pro at Spec of CP, those that license pro at Spec of INFL, and those that do not permit the licensing of pro. Second, even this offered solution fails to specify where and when subjects will be dropped. It only describes how dropping, should it occur, would be licensed.   This fails to be predictive, in contrast to OT accounts of the same phenomenon.

Grimshaw and Samek-Lodovici (1995) propose that subjects in a pro-drop language such as Italian are dropped when coreferent with the topic of the discourse. This clearly violates the principle of autonomy of syntax, but in so doing, it improves upon the previous analysis because it explains in terms of a discourse context precisely *when* a subject *must* be dropped and when a subject *must not* be dropped. It is, therefore, predictive and empirically testable. As Grimshaw and Samek-Lodovici maintain, null subjects are really null topics — an insightful claim that information structure requirements account for the missing subjects. In other words, where subjects are indexed with the discourse topic, they must be null. One reason for this may be that they are unambiguously recoverable, but regardless of the reason, the information structure requirements of the language requires the subjects to be missing, even when this creates an apparent violation of syntactic requirements (such as the unhedged version of the Extended Projection Principle).

Similar observations may also be made regarding focus. As we saw earlier, Grimshaw and Samek-Lodovici (1995) have shown that subjects which undergo what was previously believed to be free inversion are not free at all; these subjects are actually focused in the discourse. Thus, concerns for information structure relations places conditions on inversion. If focus, like topic, interfaces with syntax in significant ways to determine the acceptability of sentences, then the autonomy of syntax cannot be maintained. Recall that this was precisely the conclusion also of Costa (1997), who used Portuguese, English, Dutch, and Icelandic to illustrate differences between languages where there is a correlation between word-order and discourse functions and languages where this correlation is missing. Costa observed, and so should we, that syntactic

constraints are violated here only when they interact with discourse constraints. But the fact that these syntactic constraints are violated militates against any generativist approach in which such principles must be held inviolable. To account for apparent exceptions, other generative approaches are obligated to incorporate even greater optionality into the application of principles. Using the OT approach not only permits an account for the fact that syntactic principles are violated, but also permits us to identify the contexts where such violations occur, reducing the optionality present in the grammar.

For these reasons, OT provides an empirically cleaner, more elegant, and more predictive way of accounting for interactions between discourse and syntax. The theory is cleaner in that it does not need to feign inviolability of its grammatical principles, more elegant in that it can characterize differences in grammars as a simple reranking of the same universal constraints, and more predictive in that it envisages that the effects of a given constraint may be found independently of its being very high- or very low-ranked in a particular grammar and may specify just where we would expect to find such effects.

One final example of the interaction between discourse and syntax is that found in a recent study by Pérez-Leroux, et al. (1999) that shows how interfaces between discoursal and syntactic constraints in pro-drop languages control the distribution of null and overt pronouns. Pérez-Leroux, et al. (1999) show that in some Spanish cases syntax appears responsible for prohibiting the instantiation of a pronoun, and in other cases discoursal or contextual requirements determine the choice, as shown 2.61 and 2.62:

(2.61.) *María y Ø vamos de paseo.
Maria and   go-1pl on walk
'Maria and I go for a walk.'

(2.62.) *Ella/Ø no tiene dinero.*
         she-3sg not has money
         'She doesn't have money.'

Pérez-Leroux, et al. (1999) observe that examples such as 2.61 demonstrate that coordination and null pronouns do not mix; whereas, in examples such as 2.62, it is the need for clarity that determines whether the pronoun may be null or overt. Examples such as 2.62 are additionally constrained by the fact that old information or discourse topics spark the use of null pronouns, whereas new or contrastive information normally requires an overt pronoun.

Pérez-Leroux, et al. (1999) do not appeal to OT to resolve this interaction between discourse and syntax; however, they represent a growing line of research in the last decade that has increasingly investigated interactions of discourse grammar with syntax in such a way that challenges traditional notions of the autonomy of syntax (e.g. Vallduví 1992, Erteschik-Shir 1993, Lambrecht 1994, Grimshaw and Samek-Lodovici 1995, Samek-Lodovici 1996, Costa 1997).

This section had the singular purpose of providing a defense of my research program, whose analytical approach uses OT to explain interactions involving a discourse/syntax interface, against the critique of those who challenge such an approach as a violation of the autonomy of syntax. Although the defense I have presented shows that Newmeyer has not persuasively demonstrated that OT syntax is inherently flawed, it should still obvious that there are essential issues remain to solved. This is unsurprising, since many of these issues have not been resolved (or even addressed) by other generative approaches, despite 40+ years of research within those paradigms. OT syntax is forging a new path to address these issues, and as with most intellectual pioneers,

researchers in this field also are encountering the disparagement of cynics. Despite this, the OT theoretical framework holds enough promise to warrant its continued exploration, even in the face of hearty criticism.

2.8 Limitations of the major proposals regarding pro-drop

In this section, I briefly summarize some of the limitations of the proposals that were reviewed in this chapter, and I suggest how future research may provide some remedy for these shortcomings. The first part of this chapter focused exclusively on pro-drop. Certain themes persist in the explanations that have been given. For example, from the earliest explanations of pro-drop through some of the more recent OT accounts, a great deal of attention has been paid to the importance of inflectional agreement or identification. Whatever the mechanism used for identification, most previous research relies upon the inflectional system and agreement markers of a language for answers, either by maintaining that the verbal inflectional system must be sufficiently rich, or by arguing that it must be sufficiently uniform.

Another recurrent theme in the literature, especially the early literature, is the optional application of rules to derive the effects of pro-drop. This line of research suffers from two limitations. First, it is questionable whether the earlier proposals that appealed to complex rule-based and transformational processes can adequately address the issue of learnability in acquisition. Later parameter-setting approaches appear to address this problem, but do so only by sacrificing empirical adequacy. For example, as noted earlier, although many researchers have assumed that null subjects and free inversion are triggered by the same parameter, there are certain Italian dialects that permit

inversion, but not null subjects, and there are languages such as Portuguese, that permit

null subjects, but not inversion. To group disparate phenomena together under a single

parameter, even when those phenomena are not necessarily linked in some languages and

are acquired at different stages of a learner's development in other languages, is no more

explanatory than saying the learner once spoke a non-pro-drop language but now speaks

a pro-drop language. Such a statement is non-arguable, but also non-explanatory. If it

can be shown that the pro-drop phenomena are not a single parameter, but a fortuitous

outcome of a particular grammar, then the validity of the parameter-setting approach to

pro-drop is called into question.

Some of the earlier approaches held that clitics provide unique evidence that must

be considered by a theory of null subjects; however, the line of reasoning followed by

those who would invoke a 'clitic-parameter' should be scrutinized. Clitics do not behave

in the same way as subject pronouns. For example, clitics can not be emphasized or

coordinated, and they have a fixed order when they co-occur, all in contrast with

pronouns. Also, they are by definition unstressed, whereas pronouns such as *nosotros*

and *usted(es)* normally require stress. Furthermore, as Suñer and Lizardi (1995) note,

pronouns, unlike clitics, can be separated from the verb by negation. Once again,

applying specialized parameters to a problem does not actually solve anything; it simply

stipulates an answer.

Previous research has certainly helped clarify some of the issues involved in pro-

drop. Huang (1989) admitted discourse factors into the discussion of null subjects, an

important contribution. But Huang (1989) does not go far enough. For example, he

shows how discourse factors play a role in deciding why a null element cannot appear as

the subject of an English finite clause, but he says nothing about why Spanish finite clauses will sometimes permit null subjects and at other times prohibit them. Huang (1989) and most earlier accounts of pro-drop aim at the less ambitious task of predicting in which languages pro-drop <u>can</u> occur; but these accounts do not tell us when pro-drop <u>will</u> occur.

The Morphological Uniformity Hypothesis of Jaeggli and Safir (1989) also revealed an interesting correlation between some languages, but it does not say, for example, where or when Chinese or Spanish will choose to use null subjects. Furthermore, it does not explain why some languages with very uniform inflectional systems (e.g. German or Swedish) make little or no use of null subjects, and it fails to predict which languages will use null subjects. For example, Bani-Hassan Arabic does not uniformly mark verbal inflections, still it uses null subjects for some of its verbal forms but not others (Kenstowicz 1989). For this the MUH has no explanation. Studies by Lyons (1989), Platt (1989), and Davies (1996) provide direct empirical tests of, and counterevidence to, the MUH for adult second language acquisition.

The concerns of second language acquisition will be addressed in the next chapter, but here it may be said that if learner grammars in SLA adhere closely to UG principles, then these studies challenge Jaeggli and Safir's hypothesis. A clear prediction of the MUH would be that no languages will surface that are both [–uniform] and [+null subjects]; however, this is not supported empirically. Lyons (1989) found no significant correlation between the admissibility of null subject in a second language and knowledge of the verb agreement system. Platt (1989) showed that even when learners recognized that English verbal paradigms were not uniform, they still dropped subjects in tensed

clauses. Davies (1996) found evidence that the recognition of non-uniform verb agreement morphology on the part of his learners was neither necessary nor sufficient to determine that null subjects should be banned in English. As Davies (1996) points out, among the available conclusions are either that UG does not play a role in SLA, or the MUH must be abandoned as part of UG.

As for the more recent accounts, Radford (1997) adds feature checking to the older stories of rich agreement and identification, but it remains unclear how feature checking will resolve the pro-drop question for languages with no inflectional system, and claiming that the need to check features can explain which languages will allow pro-drop does not yet explain why pro-drop will occur for a given utterance in any language. As for Radford's (1997) example case, it should be noted that Old English had a number of properties that do make it difficult to classify it with standard pro-drop languages, undercutting part of the thrust of his argument. For example, inversion of pronominal subjects was greatly restricted in Old English (Allen, 1980), and null subjects were normally limited to expletives (Adams, 1987) and otherwise very restricted (Bresnan, 1976). These restrictions would not be predicted in the account of Radford (1997).

Speas (1997) begins well by looking at the interaction of violable constraints, but the constraints she chooses fail to account for interactions between syntax and discourse and then her account appears to fall back on the solutions (e.g. 'rich agreement') originally offered in other generative frameworks against which she is arguing. Nevertheless, Speas (1997) provides an example of how Optimality Theory could be used to approach this problem. Its fails only in that, while it does capture some generalizations

more elegantly, it does not produce a better explanation of differences between the grammatical systems of English and Spanish.

Grimshaw (1997) and Baković (1997) both contributed to a discussion of *that-trace* effects from an OT perspective, but these do not take it the additional step to show the predictions these types of constraints would make crosslinguistically. This problem is overcome in the approach of Grimshaw and Samek-Lodovic (1995), but although this study provides a theoretical analysis of differences between two language systems, it does not address developmental issues involved in the acquisition of the constraints responsible for null subjects, inversion, or *that-trace*.

It is a fair assessment of all the studies reviewed in this chapter to say that they make important contributions to linguistic theorizing concerning pro-drop, but they do not provide a developmental account of pro-drop in second language acquisition. For this reason, the next chapter reflects upon pro-drop research conducted within the field of second language acquisition.

Notes

[1] Although most of the key terms in this chapter are defined, the discussion in this chapter assumes some familiarity with generative syntax. Those requiring a broader introduction to the goals and concepts of syntactic theory may wish to consult Cook and Newson (1996), Chapters 17-25 of Radford, et al. (1999), or the first six chapters of Fromkin, et al. (2000).

[2] The concept of 'parameters' is discussed in greater detail in Section 2.3. Whether this parameter is referred to as 'pro-drop' (as was common in the early literature) or 'null subject' (as later became common), is of no significance here. Although the term pro-drop will be used throughout this dissertation, I will argue that the presence of an over-arching pro-drop 'parameter' is something of a convenient idealization that researchers have used as shorthand to group languages that behave similarly. This dissertation will argue that there is no parametric setting that yields all of the properties shown above. Rather, these properties are acquired at differing times, in a specific order, and as a result of a predictable reordering of constraints within a grammar.

[3] The concept of the 'richness' of a verbal paradigm has never been adequately defined. Presumably languages such as English and French (having only two or three inflectional distinctions for person and number) are not sufficiently 'rich', while languages that have fuller paradigms may pass some richness threshold.

[4] It has been a common premise within syntactic theory that the human cognitive system interfaces with the articulatory-perceptual system (Phonetic Form, or PF) and the conceptual-intentional system (Logical Form, or LF). The elucidation of Chomsky (1981) given here is patterned after Rizzi's discussion on the PRO' Hypothesis (1982:159-161). It should be noted that this early account uses big PRO, the phonetically null element found in the subject position of control structures, as a label for the null subject. Chomsky (1982) later treated null subjects as a separate category, pro. Others continued to conflate PRO and pro into a single empty category, accounting for differences between them by invoking independent principles and parameters (e.g. Huang 1989, Borer 1989). More recently, Alexiadou and Anagnostopoulou (1998) argue that the EPP, and consequently pro, is no longer needed under Minimalism, and Petrovitz (1999) has argued against the existence of either PRO or pro categories altogether.

[5] The details of each of these subtheories of GB (Chomsky 1982) will not receive detailed explanations here. 'Case Theory' broadly deals with the assignment of abstract Case, based on notions of 'government' found within GB theory. Under this theory, all non-empty NPs must be assigned Case. 'Theta Theory' deals with the assignment of thematic roles, requiring that arguments of verbs be assigned one (and only one) such role. 'Binding Theory' is concerned with the conditions that formally relate certain elements within a sentence. For example, as already mentioned, anaphors must be bound

and pronouns must be free in their governing categories, and referring expressions must be free.

6 Generative syntax has sometimes relied used an upside-down 'T-model' to illustrate the components of grammatical knowledge, as in Figure 2.7:

Figure 2.7 T-model of grammar

D-structure
|
movement
|
S-structure

PF component   LF component
'sounds'   'meanings'
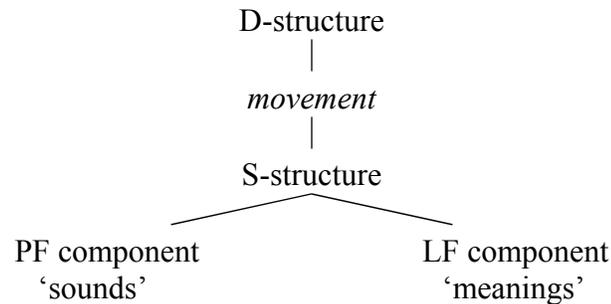
In Figure 2.7, D-structure (underlying form) is related to S-structure (form of the sentence after movement; includes traces of moved elements), and S-structure is interpreted by the two components yielding phonetic and semantic representation.

7 Suñer preferred a PRO analysis because the presence of PRO allowed parallelism between missing subject elements for both matrix and embedded clauses. Suñer's (1982) position still distinguished three types of null elements (1982:55-56, 74).

8 'Phi-features' refers to the morphosyntactic indications of person, number, gender, and case.

9 Huang (1984) showed that the problem of identification here is restricted to null subjects. Although null subjects may be regarded as pro, Huang analyzes null objects as a variable A-bar bound by an empty operator. Condition C of Binding Theory would correctly predict it to be A-free.

10 In general, Safir (1985) argues that assignment of nominative case by a subject clitic to a postverbal NP permits inversion.

11 The term 'features' has played a central role in syntactic theorizing for some time. Grammatical features included person, number and gender (also called 'phi-features'); the verbal features [± past],[± tense]; and the binding features [± anaphoric] and [± pronominal] that stem from Chomsky (1981).

12 'Procrastinate' is a principle favoring covert syntax over overt syntax, when there is a choice. The concept also involves the Minimalist idea of 'Spell-Out', which tells the grammar to move to create the Phonetic Form. Movement that takes place after Spell-Out is not as 'costly' as before Spell-out, because it does not have to pied-pipe phonological features (which are stripped away at Spell-Out), but delaying movement until after Spell-

Out can only take place with weak features. If strong features are involved, these are thought to be uninterpretable at PF, and therefore they must undergo deletion prior to Spell-Out.

[13] This dissertation assumes that part of the universal construction of GEN is that it follows X-bar principles in generating outputs. This critically assumes that all languages obey X-bar principles, which is an empirical question for further research.

[14] Grimshaw (1997:414) notes an exception for relative clauses, where 'that' is required to be present when it is the highest subject that is extracted.

[15] A more detailed discussion of the problem of language acquisition is found in the next chapter.

[16] Even the most ardent opponents of OT are often circumspect about attacking the value of OT for phonological research, perhaps due to the extraordinary and sweeping impact OT has had for that field.

[17] I have used the term 'formalist' here rather than 'generativist' to not place OT outside of the generativist approaches. It is interesting to note that there are OT researchers in both of the camps contrasted here; for example, Haspelmath (1999) and Aissen (1999) use functionalist approaches within an OT framework, but other researchers such as Grimshaw (1997), Samek-Lodovici (1996), or Costa (1997) consider interactions between different levels of grammar without resorting to functional explanations.

[18] Data complexity refers to the amount of data needed for an algorithm to ensure that it learns a correct grammar. Since the Constraint Demotion Algorithm (Tesar and Smolensky 2000) operates on informative data pairs, the maximal distance between an initial hierarchy and a target hierarchy is $N(N\text{-}1)/2$, where N is the total number of constraints.

[19] Exactly what constitutes input in OT has been a matter of some debate. This dissertation follows the view of Grimshaw and Samek-Lodovici (1995:590), who regard input as 'a lexical head and a mapping of its argument structure into other lexical heads, plus a tense specification'.

[20] There are contexts where the complementizer is obligatorily present or absent, and Baković accounts for this as well, but for the purposes of our discussion here, we will focus on the optionality issue.

[21] Some researchers also include other elements (e.g. Baković's [SUB]) as part of the input.

[22] Legendre (1999) showed that in French, inversion indicates that the subject is 'identifiable' whereas non-inversion indicates that the subject represents active old information. Legendre thus uses information structure terms such as those discussed in

Lambrecht (1994) to construct a constraint that aligns non-active NPs with the right edge of V^0. This constraint, in competition with a constraint requiring subjects to be realized in Spec IP, yields the correct empirical results.

[23] For example, left dislocation can be used to trigger a partially ordered set (poset), to aid in discourse processing or to permit island violations, and topicalization either triggers poset inferences or simply serves in the traditional topic-marking role.