

Minimal Computation: Derivation of Syntactic Structures

by

Charles D. Yang

Submitted to the Department of
Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of
Master of Science

at the

Massachusetts Institute of Technology

May, 1997

©1997 Massachusetts Institute of Technology. All Rights Reserved.

Signature of Author _____
Department of Electrical Engineering and Computer Science
May 23, 1997

Certified by _____
Robert C. Berwick
Professor of Computer Science and Engineering
Thesis Supervisor

Accepted by _____
A. C. Smith
Chair, Department Committee on Graduate Students

JUL 24 1997

Eng.

Minimal Computation: Derivation of Syntactic Structures

by

Charles D. Yang

Submitted to the Department of Electrical Engineering and
Computer Science on May 23, 1997 in partial fulfillment of the requirements
for the degree of Master of Science

Abstract

The Minimalist Program aims to eliminate rules and structures that are not absolutely necessary for linguistic description and explanation. In this framework, linguistic expressions are generated by optimally efficient derivations that must satisfy the conditions that hold on the interface levels, the only levels of linguistic representation. The interface levels provide instructions to two types of performance systems, articulatory-perceptual and conceptual-intentional. All syntactic conditions, then, express properties of these interface levels, reflecting the interpretive requirements of language and keeping to very restricted conceptual resources.

This thesis tries to substantiate these claims. Conceptual motivations to reduce complexity lead to a set of necessary conditions on the computational system. The challenge is to show that those conditions are sufficient for the coverage of empirical facts. The point of departure is the expletive construction and its relation to economy. It is proposed that only features necessary for local checking are attracted in the computation. The traditional notion of EPP is reduced to the interactions between the Case system and a parametric PF condition. Many desirable consequences follow including a unification of null and displaced subjects. In the proposed system of grammar, language variations are literally determined and expressed at interface levels, with computational complexity reduced to minimum.

Thesis Supervisor: Robert C. Berwick

Title: Professor of Computer Science and Engineering

Acknowledgments

First of all, I wish to thank my teachers: To Bob Berwick, who got me started and gave me support and guidance that went much beyond this thesis; To Noam Chomsky, for many aspiring meetings, comments, letters, and e-mails, all under his overwhelming schedule. This report started out almost two years ago when I tried to do a decent term paper for Noam's class.

Thanks extend to many colleagues and friends. Howard Lasnik read an earlier draft and made many sharp and detailed comments. He showed me what serious scholarship is all about. Sam Epstein got me (really) interested in syntax with his "Un-Principled" paper, and has kept me interested with enthusiasm and encouragement. Chris Collins made his (then unpublished) work available, therefore provided me with concrete problems to work on. Although I don't agree with Chris on certain issues, it is clear that without his work, this thesis would never be possible. Special thanks to Edith Kaan, Julie Legate and Carolyn Smallwood, who listened to many of my ideas and made many useful suggestions, particularly some last-minute heroics for turning this thing into English. I'm also grateful to Youngjun Jang, Bob Kuhns, Shalom Lappin, Alec Marantz, Martha McGinnis, Colin Phillips, Jeannette Schaeffer, and Ed Stabler for helpful discussions and comments at various stages of this work.

Finally, I'd like to thank the past and present residents at the Church St. Inn, and friends in the AI Lab and elsewhere, for good times and putting up with my perverted queries on expletives.

This work is supported by a National Science Foundation graduate fellowship.

Contents

1	The Minimalist Approach	7
2	The Optimality Hypothesis	10
2.1	Whither Economy?	10
2.2	Complexity and Interface	12
2.2.1	Global and Local Economy	13
2.2.2	On the Role of Interpretability	16
2.3	Minimal Computation	20
2.4	The Economy Principles	22
3	Expletives and Economy	26
3.1	Expletives and the MPLT Analysis	27
3.2	Objections to the MPLT analysis	31
3.2.1	Objection # 1: Optionality	31
3.2.2	Objection # 2: Complexity	37
4	T	44
4.1	Minimal Feature Attraction	45
4.2	The nature of T	49
4.3	Expletives Revisited	52
5	Null and Displaced Subjects	58
5.1	Null Subjects	58
5.2	Displaced Subjects	64

5.2.1	Locative Inversion	64
5.2.2	Other displaced subjects	68
6	Other Minimal Properties	72
6.1	Interpretability and Economy	72
6.2	Cyclicity and Generalized Transformations	74
7	Conclusions and Speculations	78
7.1	Perfect Language	78
7.2	Imperfect Use	81

Chapter 1

The Minimalist Approach

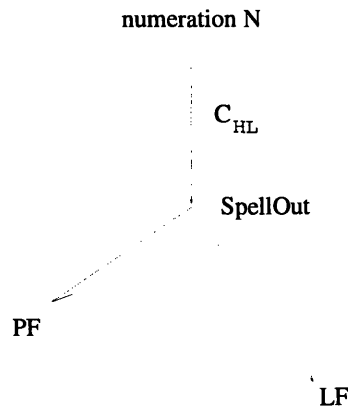
Linguistics is the study of Form and Meaning and their relations. Situated in a broad theory of mind, a linguistic theory aims to characterize and explain the properties at the perceptual (Form) and conceptual (Meaning) interfaces, and the rules and representations employed to relate them.

The Minimalist Program (Chomsky 1995) outlines a research framework rooted in this “virtual conceptual necessity”. It proposes that the essential properties of human language, including cross-linguistic variations, are largely determined by the conditions at the external interfaces: the Phonetic Form (PF) and Logical Form (LF), respectively. These conditions impose constraints on the possible structures of human language. A PF condition says, for example, that a structure to be pronounced cannot have a stressed consonant or any other physically / phonologically impossible sound patterns. An LF condition would require that natural language does not allow vacuous quantifiers, as in “ \star *What John likes Mary*”, contrary to some formal systems where $\exists x(2 + 2 = 4)$ is technically admissible. Following Chomsky (1995), we call such interface constraints altogether the *Bare Output Conditions*.

The computational system of human language C_{HL} is a generative procedure that composes linguistic structures to satisfy the Bare Output Conditions. Specifically, lexical items in a *Numeration* N are assembled by structure-building operations, the *Generalized Transformations* (GT) that apply recursively, in the Humboldtian sense of “infinite use of finite means”. The computation branches at the point of

Spell-Out to yield a pair of representations $\langle \pi, \lambda \rangle$ for the PF and LF interfaces respectively (figure 1). It is conjectured that the C_{HL} exhibits a certain degree of “optimality” or “elegance”: the computation is governed by economy principles that avoid superfluous operations and extraneous representations. Under this view, the C_{HL} is something like a “perfect solution” to the Bare Output Conditions, understood as computational effort in a sense to be made precise. Call this perfection conjecture the *Optimality Hypothesis*.

(1)



The Minimalist Program.

I am interested in formulating and understanding the following questions in this framework:

- (2) (a) To what extent are language variations determined at the external interfaces?
- (b) Given a concrete computational measure of economy, how “perfect” is the C_{HL} that maps N to $\langle \pi, \lambda \rangle$ to satisfy the Bare Output Conditions?

Some qualifications. To even start considering the problems in (2), one must assume that computational complexity or efficiency plays a role in the evaluation of linguistic theories, a strong assumption that is perhaps dubious for cognitive systems

in general. It is also important that this complexity analysis be understood as intrinsic properties of the I-language (Chomsky 1986), not to be confused with parsing efficiency and other E-language notions. Nor should it be equated to a production or processing theory of performance that incorporates the knowledge of language. A theory of the C_{HL} is an abstract system which specifies the state of knowledge that the speaker of language possesses, and the conditions the yet unknown mechanisms must meet.

I will pursue a theory that has the character of “minimal computation”. This theory makes use of some conceptually simple and natural principles to minimize computational cost of the syntactic derivation. Conceptual motivations are given first, followed by empirical problems and analysis.

This thesis is organized as follows. Chapter 2 discusses the role of economy and complexity, and outlines the scope of a minimalist theory to the extent consistent with the Optimality Hypothesis. It is argued that the computational system should be constrained to a very narrow range and economy evaluation strictly local. Chapter 3 starts the empirical study. The point of departure is the expletive construction. A number of proposals are reviewed and their conceptual and empirical problems are pointed out. A solution is proposed in Chapter 4, grounded in the feature checking and attraction framework and general economy conditions. Some predictions and consequences are discussed in Chapter 5, where I try to unify expletive constructions, locative and quotative inversions, and certain aspects of null subject languages. Chapter 6 briefly discusses two additional problems, the treatments of interpretability and the strict cyclicity condition, again with the aim to eliminate computational complexity while retaining empirical coverage. Chapter 7 concludes with some general remarks on the complexity of C_{HL} and speculations on the computational structure of human language.

Chapter 2

The Optimality Hypothesis

2.1 Whither Economy?

There is of course no *a priori* reason that the C_{HL} should be computationally “optimal”. In general, precision and efficiency are not known to be defining characters of cognitive systems. Nonetheless, recent developments in syntactic research seem to suggest that some unifying principles with a flavor of economy do cut across languages and constructions, and have opened up new and deeper topics of inquiries. Therefore, we adopt the Optimality Hypothesis as a theoretical conjecture (that must be ultimately tested out empirically) and explore its consequences.

Note that the Optimality Hypothesis does not arise for theories that do not or cannot make computational complexity claims. Broadly, such theories can be called **representational theories**. A representational theory typically consists of a set of conditions or constraints that is imposed upon some (candidate) structures and hence determines the legitimacy of these structures. An example is the Optimality Theory (Prince and Smolensky 1993). The Optimality Theory assumes that constraints are violable and ranked. All candidate structures are produced by a generative function GEN (presumably a deterministic procedure). The structure that least violates the constraints with respect to their rankings is deemed the optimal structure. No specific claims have been made about the mechanical complexity of GEN, and it is not clear how to do so either (in an intuitive and natural way). Another example is Brody’s

(1995) approach to the Minimalist Program – again, Brody is not concerned with the generation of the representational structures (at his Lexico-Logical Form) but instead with the conditions they must observe. On the other hand, a derivational approach takes the recursive procedure literally: “it forms expressions step-by-step by applying its operations to the pools of features” (Chomsky 1996). Such a theory, by definition, must address how structures are built by the computational procedure, so that claims like the Optimality Hypothesis can be formulated rather directly. I **do** want to emphasize that it is a separate and ultimately empirical question what the “nature” of human language **is**, be it derivational or representational. Both approaches must describe and explain the same set of linguistic data. The “ease” of formulating computational statements like the Optimality Hypothesis does not *a priori* grant any advantage to the derivational approach, since the role of complexity in the study of cognitive systems is still largely unknown, and often misunderstood, such as the “psychological reality”, parsing efficiency, and others. For more discussions of this, see Yang (1996).

I will nonetheless assume that C_{HL} is indeed a derivational system, because interesting problems such as the Optimality Hypothesis are opened up for investigation. I find it convenient and perhaps also necessary to cast a grammatical theory in a rigorous formalism for which proposals can be concretely formulated and examined: a mechanical system that blindly carries out instructions as specified by the grammar. This task is difficult to accomplish for a theory that does not spell out mechanisms of the generative procedure. Furthermore, there seems to be some results in favor of the derivational approach. For instance, Epstein (1995)¹ proposes a derivational formulation of the notion c-command, which explains why c-command is **stipulated** as such in representational theories, e.g. the first-branching node definition of Reinhart (1979). Under the derivational framework, the grammatical system consists of a small number of construction-independent rules, the Generalized Transformations. The application of these rules (not resulting candidate structures, as in representa-

¹See also Groat (1995), Chomsky (1996), Abraham et al (1996), Kitahara (1997), Epstein et al (forthcoming) for extended discussions.

tional theories) is directly constrained by economy principles. For instance, when an element undergoes movement, principles such as the Minimal Link Condition (Chomsky and Lasnik 1993) dictate that it cannot move further than its closest landing site.

Once the derivational approach is adopted, two technical problems immediately arise. First, we must show that the C_{HL} derives only structures allowed by the interfaces – certain but not arbitrary structures are part of human language. Second, we must show that in the process of derivation, the C_{HL} demonstrates computational optimality: for instance, cheaper derivations are favored over more expensive ones, given a proper metric for cost evaluation. This is forced by the Optimality Hypothesis. These are the two principal motivations for our discussion. Let's consider them in turn.

2.2 Complexity and Interface

First and foremost, an adequate Minimalist theory should admit only the derivations that ultimately satisfy the interface conditions. In other words, the C_{HL} should not over-generate or under-generate. It cannot over-generate by admitting divergent derivations and cannot under-generate by failing to admit convergent derivations. Some concrete definitions (3), following Chomsky (1993, 1995), Collins (1994, 1997), and others:

- (3) Given a numeration N of lexical items, a derivation D outputs a PF and LF representation pair $\langle \pi, \lambda \rangle$. D is:
 - (a) *Legitimate*: if all the morphological features (e.g. Case, [-wh]) in $\langle \pi, \lambda \rangle$ have been eliminated or satisfied at PF and LF.
 - (b) *Interpretable*:² if λ meets the Theta Criterion and Full Interpretation (FI) at LF.
 - (c) *Convergent*: both legitimate and interpretable.

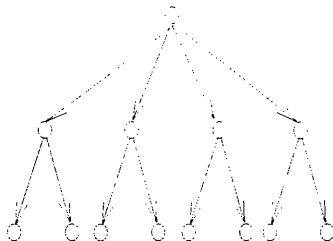
²This is relevant to the treatment of the Theta Criterion and complexity considerations. See section 2.2.2.

It is convenient to view the derivation as a search problem, typical in artificial intelligence, e.g. Winston (1994). A derivation D corresponds to a path starting from N to the structure S that D generates. The target of the search is the convergent derivation D_{con} which satisfies the interface conditions BOC. Economy principles, for example, measure the number of steps in a derivation, the distance of a moved element from its base position, so on and so forth (more in 2.3), and select operations with cheaper costs. The Optimality Hypothesis is attained via economy principles, which constrain the range and possibilities of derivations.

2.2.1 Global and Local Economy

There are many ways to construct a theory of the C_{HL} , and they have distinct computational properties. An obvious one can be called the **strongly global theory**:³

- (4) For a given N , compute all the derivations and select the most economical one.



Global Computation

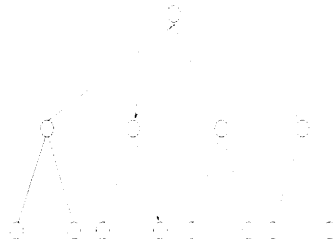
Such a theory literally enumerates all the derivations from N . By comparing the costs associated with them, the optimal derivation can be determined. The search procedure corresponds to a Depth (or Breadth) First Search (Winston 1994: Chapter 4), for which the entire search space is potentially expanded out. The success of the search is guaranteed, that is, the most economical derivation D_{eco} will be found, since all possible derivations are explicitly computed and compared. However, the induced

³Johnson and Lappin (1996) examined a theory similar to this and correctly pointed out some complexity problems. I will return to their analysis as we go along.

computational complexity is immense, as the resource requirements on space and time (metaphorically, the size of the search tree) grow exponentially. The arrows in (4) indicate the potential search space the global theory has to traverse. The computation is **global** because economy conditions are invoked only on global structures, i.e. complete derivations. This aspect of the global theory is not unlike the representational theories discussed in section 2.1, that constraints (economy metrics, the interface conditions, etc) are invoked only when the derivations are completed and candidate structures have been fully built. In other words, economy doesn't do the job to trim down the search (derivation) space. Strongly global computation clearly runs against the Optimality Hypothesis and must be rejected on conceptual grounds.

By contrast, another theory can be called a **strongly local theory**:

- (5) At each stage of the derivation, pursue only the most economical operation and abandon all the rest.



Local Computation

In (5), economy principles are formulated and applied strictly locally. Such a theory corresponds to the so-called Best First Search (Winston 1994: chapter 5). The computational complexity in a local theory is radically reduced to minimum – in fact, linear to the depth (height) of the search tree. For an optimal solution, the computation has to make local use of the interface conditions **in the course of the derivation**. The empirical question is, naturally, if this local theory can find the optimal derivation, a goal that is attainable under the global theory but only at a formidable cost.

With complexity considerations as our conceptual guideline, we want to find out what kinds of economy principles would constitute an optimal theory. It has been

noted ever since the onset of the Minimalist approach (Chomsky 1988/1991) that some proposed principles do introduce enormous complexity – seemingly paradoxical given the Optimality Hypothesis that the C_{HL} is a simple computational system. There are a number of possibilities to reconcile this paradox. Chomsky notes that the C_{HL} might have access to some “computational tricks that will overcome the problem of intractability” (1988/1991: p49). For instance, since Case-marked elements cannot undergo A-movement, the computation does not have to consider movement of such elements. In this sense, it is logically possible to allow an computationally intractable system, as long as we have some handy tricks to overcome intractability. However, as I will argue in section 2.3, even fragments of intractability can undermine the Optimality Hypothesis altogether, and it is unclear how to design computational tricks in a principled and systematic way.⁴ On the other hand, it is perhaps good research strategy to push conceptual assumptions to extremes, to test out their validities or to reveal potential problems. Therefore, my solution to this paradox is to abandon in entirety the intractable principles, and hold the Optimality Hypothesis in the strongest possible form:

- (6) (a) Reject a theory for which the complexity is intractable
- (b) Admit only a theory for which the complexity is minimal⁵

(6) expresses a tension between our conceptual motivations and the empirical burdens of descriptive adequacy. We are driven to construct a theory of the C_{HL} that demonstrates computational optimality but we must show it is (at least) descriptively adequate; otherwise the simplicity or “elegance” of such a theory would be immediately surmised. It is this tension that drives the discussions in this thesis. I will argue that such a theory is perhaps attainable.

⁴See section 6.1 for a principled “trick” for the treatment of Theta Criterion, based on Hale and Keyser’s (1993) configurational approach to argument structures.

⁵Or at least tractable, which means that the time/space resource requirement of a derivation is polynomially bounded by the problem size – the standard definition of a computationally simple system.

2.2.2 On the Role of Interpretability

Before we investigate specific properties of the C_{HL} , consider another conceptual issue. It has to do with the role of Interpretability (defined in (3)) in the computational system. Chomsky notes that as an autonomous system, “derivations are driven by the narrow mechanical requirement of feature checking only, not by a ‘search for interpretability’ or the like” (1993: p 33). It is crucial to understand this remark properly – it **cannot** mean that we should somehow ignore interpretability altogether, or cast it outside of the local computation system. Instead, both interpretability and legitimacy (that is, convergence) must be guaranteed by local economy.

As a simplification, let’s take interpretability to be the proper satisfaction of the θ -criterion. Chris Collins (1997: p71) argues that violation of the θ -criterion, hence interpretability, does not constitute non-convergence. Collins notes the contrast between the following pair (7):

- (7) (a) ★ John seems that he is nice.
(b) ★ Arrived John.

(7a) has all morphological features properly checked, but contains a violation of the θ -criterion because *John* has no θ -role. (7b), on the other hand, satisfies the θ -Criterion (*arrive* is unaccusative) but fails to check off the Case feature of the finite T. Collins argues that there is a qualitative difference between this pair and concludes, tentatively as he himself notes, that the ungrammaticality of the pair should not be compared, because they are violations of different nature.

I have no interesting analysis for the contrast in (7). The relationship between the θ -Criterion and convergence is ultimately an empirical question. But there is an important conceptual consequence that follows from Collins’s suggestion.⁶ Briefly, for Collins (1997) as well as Chomsky (1993, 1995), legitimate but **uninterpretable** derivations can be generated, at least at one point of the derivation; presumably at

⁶The impact can be seen in the status of Procrastinate and some analyses of expletive constructions. I will discuss the technical details in 4.2.2.

LF, they can be ruled out instantly.⁷ In this sense, their approaches provide the interfaces (LF) a set of “candidate” solutions, without reference to interpretability (the θ -criterion). It is true for Collins (1997) that each step of the derivation observes strict local economy. This seems to be compatible with Chomsky’s remark quoted in the beginning of this section: a computational procedure that just checks off morphological features, and is not driven by the “search for interpretability”. However, this approach is not compatible with the Optimality Hypothesis, though at the interfaces, uninterpretable derivations can be ruled out without much computational cost. This incompatibility directly follows from computational complexity theories.

The basic claim is this. A computational system S that provides **candidate** solutions, no matter how efficiently the candidates are generated, or how efficiently they can be verified, says **nothing** about the intrinsic complexity of S to give the **right** solutions. In fact, it is well-known that “easy-verifiability” is a useful (but neither sufficient nor necessary) diagnosis for computationally **hard** problems. Take a classic example, the 3-Satisfiability (3SAT) problem⁸ (Gary and Johnson 1979):

- (8) Collection $C = \{c_1, c_2, \dots, c_m\}$ of clauses on a finite set U of variables such that $|c_i| = 3$ for $1 \leq i \leq m$.

Question: Is there a truth assignment for U that satisfies all the clauses in C ?

When a solution is given, it’s trivial to determine whether it is correct – simply plug in the variable values. It is also trivial to generate a **candidate** solution: one can make a ‘guess’ or enumerate all the candidates exhaustively and then verify their correctness trivially. To claim that a problem is inherently simple (polynomial to its size), efficiency for candidate generation / verification alone is not enough. The problem must be easy to **solve** – either give the **right** solution, or signal that no solution exists – and do so efficiently. For the 3SAT problem, the computational

⁷We can simply count the number of unchecked morphological features for legitimacy, and examine theta-role bearing chains for interpretability, so on and so forth. Legitimacy, interpretability and convergence are defined in (3).

⁸See Barton, Berwick and Ristad (1987) for an introduction to computational complexity theory and some applications in linguistic analysis. I am glossing over much technical details here.

procedure must either give an assignment of variable values to satisfy the truth condition, or give a negative answer, “Sorry, no such assignment exists”. In either case, the computational complexity must be tractable to qualify as “easy” or “efficient”. Unfortunately, problems such as the 3SAT are notoriously hard, and no known efficient solution exists. It is generally believed that no efficient solution is likely to exist. Otherwise, we would have a negative answer to the conjecture $P = NP$, perhaps the most intriguing and challenging problem in modern mathematics. It would imply that many extremely hard problems that have been studied intensively but to no avail would suddenly become extremely easy: certainly very counter-intuitive, and almost too good to be true.⁹

Back to linguistics. The Optimality Hypothesis claims that the C_{HL} is an optimal system, a perfect solution to the Bare Output Conditions. Now we must show its computational complexity is fundamentally **different** from something like the 3SAT problem, a known tough cookie. Specifically, we must show that C_{HL} gives **right** solutions efficiently, not just easily-verifiable candidates. Whatever the proper role of the θ -Criterion turns out to be, a θ -violation is still a violation, an illegitimate (divergent) structure that is not admissible at the interfaces. A proper theory should somehow rule it out – and we cannot postpone the problem to the interface because then we would be introducing candidate solutions, not the right solutions. We must show that divergent derivations can **never** be produced, because once they are, the Optimality Hypothesis is already diminished. This is the scope for the optimal theory that I have in mind. It seems that departure from this immediately violates the the Optimality Hypothesis, our working assumption, and *a posteriori* must be rejected.

Chomsky’s remark on interpretability still holds, but we must understand it correctly with respect to computational complexity. First of all, the computation must ensure interpretability, though it still performs mechanical feature checking. Also, as will become clear in section 2.3, economy conditions by which the computation operates cannot be overridden by convergence hence interpretability. Nor can econ-

⁹Or too bad to be true, for cryptographers, who rely on the **belief** (not proof) that problems like prime factoring are hard not because we are too stupid to come up with efficient solutions, but because they are **intrinsically** hard and no efficient solution can possibly exist.

omy choose among convergent (hence interpretable) derivations (see also page 25). This is because once an uninterpretable thus divergent derivation is generated when a convergent one actually exists, the computation has to backtrack to pursue alternative derivations – which introduces complexity. Instead, economy should **ensure** convergence, when a convergent derivation does exist.¹⁰ The system is still something like a “narrow mechanical” system that just blindly satisfies local conditions to build syntactic structures, without “the search for interpretability”; interpretability follows.

More concretely, I propose a Minimal Computation Theory, which maximally reduces complexity and meanwhile ensures convergence:

(9) MCT: For a given numeration N ,

if a convergent derivation D_{con} exists for N

then for the most economical derivation D_{eco} , $D_{eco} = D_{con}$

else (no D_{con} exists for N)

if MCT still yields derivation D_{eco} , satisfying economy

then D_{con} must crash at LF or PF

otherwise

MCT must have crashed at an intermediate stage of the derivation

Here “economy” refers to the kind of radical local economy as outlined in (6). The MCT should not over-generate or under-generate. The interface conditions, including interpretability should be completely localized, by the economy principles which apply locally to ensure computational efficiency. The problem imposed by the interfaces is “solved” by the C_{HL} , with minimal computational effort. It is obvious that the empirical burden is enormous; but it is also obvious that the MCT, if true, reveals some interesting and striking properties of human language (see Chapter 7).

¹⁰There is no false alarm: for an numeration N , if local economy fails to find any derivation, no convergent derivation can exist to start with.

2.3 Minimal Computation

With the MCT (9) in mind, let's consider what a "perfect" theory should look like, where perfection is understood as computational complexity (simplicity). Let's first consider some general computational properties of the derivational system and study what kinds of principles will make the computation efficient. Some general questions:

- (10) (a) Can the C_{HL} allow multiple derivations to proceed simultaneously, and/or perform trans-derivational comparisons?
- (b) Can the C_{HL} "look ahead" or backtrack, along the paths specified in the derivation search space?

Simply put, the answer to these questions is No. Suppose (10a) were true. Suppose, for sake of argument, that in the derivation of the simple declarative "John left", two derivations are allowed to co-exist and co-proceed. Assume also that these two derivations are "distinguished", e.g. one observes the θ -criterion and the other doesn't, when the derivation reaches LF. Now consider (11):

- (11) [Carol left] before [Cindy left] after [Stephanie left] because [I left].

Each bracketed clause results in two derivations to proceed simultaneously. Therefore, (11) introduces $2^4 = 16$ derivations to co-exist. It is clear that the growth is exponential and thus computationally intractable.¹¹ (11) demands that only *one* derivation be allowed **at any stage** of the derivation.¹² In the optimal theory, (10a) must be false.¹³

¹¹Note even if the C_{HL} has a (finite) number of "processors" to allow multiple derivations in parallel, the fundamental problem of intractability still cannot be overcome, for the growth of complexity is exponential and the speedup by parallel computation is only linear to the number of processors.

¹²See also the motivation for the MCT in section 2.2.2. Suppose two derivations are allowed for each clause in (11), but only one is convergent, the other is legitimate but not interpretable. It is clear that combinatorial explosion requires an exponential number of derivations to be considered, though each of them is easy to "check" at LF. In general, we cannot afford to choose among multiple derivations, making references to non-local stages, including LF. Local economy must choose the cheapest derivation right away.

¹³Some complications need to be clarified. By "optimal derivation", I mean the optimal derivation for a **single** numeration. Consider the numeration of unordered lexical items {John, Mary, likes}. Technically, there are at least two convergent derivations for this set of lexical items: a simple declar-

Similarly, (10b) cannot be true either. Otherwise, the computation must “memo-
rize” the paths that the derivation has traversed in order to lookahead or backtrack.
This requirement, again, leads to exponential growth in complexity. In the optimal
theory, computational decisions should be local– without reference to other stages of
the derivation, and also decisive – once an option is ruled out, it’s out for good. There
is no place for errors or misses.

Examples such as (11) put very strict conditions on possible principles allowed in
the Minimal Computation Theory. No global computation is allowed **at all**, for even
fragments of it could result in combinatorial explosion. To maintain the Optimality
Hypothesis, the C_{HL} should make decisions exclusively locally, and the computational
complexity is reduced to minimum:

(12) Minimal Computation Guidelines

- (a) no reference to other stages (no lookahead, no backtrack)
- (b) no parallel derivations or trans-derivational comparisons.

More formally, consider the derivation as a transitional process that moves from state
to state. The most economical derivation (*eco* in the superscript) takes the following
form:

$$(13) S_0^{eco} \rightarrow S_1^{eco} \rightarrow \dots \rightarrow S_{n-1}^{eco} \rightarrow S_n^{eco}$$

At the state S_i , there are m operations OP_i^j , $j = 1, 2, \dots, m$ that transfer the deriva-
tion from S_i to S_{i+1}^j . These m operations constitute a **local** reference set in the sense
of Chomsky (1995: p227-228). The operation OP_i^{eco} is the cheapest, determined by

ative “*Mary likes John*”, and a topicalized “*John, Mary likes*”, leading to two distinct LF objects,
though superficially, they share the same set of phonetic features. Presumably, this distinction must
be specified in the numeration. That is, *Mary likes John* and *John, Mary likes* are derived from
two **distinct** numerations (for the latter, there could be a feature, say, [-Topic] and hence do not
compete with each other.

Similarly, the so-called “optional” operations, such as Icelandic object shift (Bobaljik and Jonas
1996) and scrambling (Fukui 1993), lead to different PF and LF objects. Hence they are products
of different numerations with different feature specifications, hence not truly optional, though they
contain the same set of “words”, phonetically pronounced. In their derivations, however, feature
checking must also observe strict local economy.

economy conditions for which $\forall j, j \neq eco, c(OP_i^{eco}) < c(OP_i^j)$, where c is the cost function associated with each operation,¹⁴ computed at S_i without making reference to any other state, satisfying local economy (6) and minimal computation (12). The derivation D_{eco} is the sequence of states S_i^{eco} with $i = 1, 2, \dots, n$, which satisfies local economy, and the Spell-Out state S_n^{eco} yields a structure pair $\langle \pi, \lambda \rangle$ at PF and LF respectively. In the Minimal Computation Theory (9), if D_{con} exists, $D_{eco} = D_{con}$. Local economy provides an efficient and also “correct” (convergent) solution to the Bare Output Conditions.

2.4 The Economy Principles

Given the conceptual background for the MCT, let’s analyze some economy principles that have been proposed in the literature and their computational complexities. Principles that involve global computation are rejected, following the conceptual arguments given in sections 2.1-3. Only strictly local economy conditions are preserved. The rest of the paper tests them out on some empirical problems.

Consider first a version of Last Resort,¹⁵ sometimes called “Have an Effect on Output” (Chomsky 1995: p294):

(14) Last Resort: do X only if X is a necessary step to check off some feature(s).

(14) is plainly a global condition, under which we might have the following scenario:

(15) (a) do X , but X doesn’t check off any feature

(b) ...

(c) ...

(d) do Y , which checks off a feature F but cannot do so unless X has occurred

¹⁴It is logically possible that two or more operations end up with a “tie” in terms of their associate costs and the computation would be forced to proceed in parallel. In the MCT proposed here, this situation cannot exist, for it directly introduces intractability. So far I have not seen evidence alluding to this possibility. However, if it did, the theory proposed here would face serious challenges.

¹⁵Its antecedent, “Greed” (Chomsky 1993, 1994), has been shown to be problematic (Lasnik 1995, Collins 1997), so we will not consider it here.

In (15), X is like a “prelude” to Y . The application of X satisfies (14) since X is a necessary yet indirect step to check off F at the stage Y . The computation, however, must look ahead to determine whether X is a necessary but indirect step for feature checking. Complexity considerations (12) suggest that computational decisions be made locally, without reference to other stages of the derivation. We therefore assume a local version of Last Resort as in (16), taking a narrower and stronger form of (14):

(16) Economy Principle I

Last Resort: Do X only if some feature is checked off as an immediate result of X .

(16) is essentially Chomsky’s (1995: p257, 20a,b) and Collins’s proposal (1996: p9). The computation is extremely short-sighted: don’t do anything unless it has an immediately impact. The situation in (15) must be either factually nonexistent, or somehow captured in the local theory.

Consider next the principle Procrastinate (Chomsky 1995: p 198):

(17) LF movement is cheaper than overt movement, and hence preferred.

Under (17), it is possible for the computation to compare derivations and their associated costs. For example, one with overt movement *before* Spell-Out is considered more expensive than another that moves covertly after Spell-Out. When referenced in the derivation, this necessarily entails global computation. First, it involves lookahead like (15): the computation doesn’t “know”, in advance, whether a particular operation should be delayed for a cheaper cost, because movement in overt syntax might well be mandatory (instead of optional) for convergence. This violates (12a). Second, if it keeps multiple derivations in “memory” for comparison, e.g. one with overt movement and the other with LF covert movement. Hence multiple derivations co-proceed, violating (12b). Therefore, (17) introduces global economy and intractability and is thus undesirable. Other formulations, e.g. that Procrastinate can be violated for convergence (Chomsky 1995), have the same flavor – LF conditions are invoked to check for convergence. They have no place in the Minimal

Computation Theory proposed here.¹⁶ In general, optionality of operations introduces enormous complexity, as noted by Chomsky (1988/1991: page 45). Here we will ban optionality systematically. The effect of this can be seen in the treatment of nominative Case checking in Chapter 4.

To set stage for later discussion, I will argue that there is indeed a kind of “invisible” movement that happens **before** Spell-Out in “overt syntax” though its effect is not necessarily manifested at PF. It refers to formal (perhaps also semantic) feature movements without pied-piping PF features. This is elaborated in Chapter 4. For sake of clarity, I will call this *Feature Movement*, applicable before and at LF, to be distinguished from *Covert Movement* that occurs exclusively at LF.¹⁷ No operation should be optional, regardless of the linguistic level at which it applies: the computational system is not free to choose or delay. Instead, whether an operation takes place before or at LF should be specified by parametric values and perhaps ultimately, interface conditions. An example is the verb movement contrast in English vs. French, following Emonds (1978), Pollock (1989), among others. Computationally, they do not enter into economy competition. Derivation is always deterministic, and that is that.

Another economy principle I will adopt is that the computation carries just enough features to satisfy feature checking. This is fundamentally different from “move just enough features for **convergence**” (Chomsky 1995: p262). “Convergence” makes reference to non-local stages (LF) and violates the computational guideline (12). The intuition is that, don’t move more than necessary. Of particular relevance here is the movement of PF features (overt categories). Presumably, whether an entire category moves is determined by parametric PF strength. Therefore, when feature movement is possible, whole category should not move, unless forced by PF conditions that are readily accessible in the derivation. Along with the Minimal Link Condition (Chomsky 1995: 311), we lump together operations that have a flavor of “as cheap as possible” as the Minimality Condition, following a term by (Collins 1997: p9) but

¹⁶A (strong) expectation of this is, for a given numeration, there cannot be more than one derivations (paths) that are equally economical leading to a single $\langle \pi, \lambda \rangle$ pair.

¹⁷Perhaps all LF movements are in fact movements of features before Spell-Out; see Chapter 4.

generalizing it to include Feature Movement:

(18) Economy Principle II

Minimality: An operation *OP* may apply only if there is no cheaper operation *OP'*.

The cost is determined by the distance of movement, the number of features that move, and others. It is important that evaluation of cost must be computed locally as well.

The principle Fewest Steps, sometimes called Shortest Derivation Requirement (SDR), has no place in the local theory because it involves explicit computation and comparisons of multiple derivations. Analyses that rely on the SDR, such as Kitahara's deduction of cyclicity (1995), should be recast in the local theory. See section 6.2.

Finally, I adopt Chomsky's suggestion that feature checking should not delay (1995: p233):

(19) Economy Principle III

As Soon As Possible (ASAP): check features immediately when possible, don't delay.

The conceptual motivation is clear: if it did, the computation might have to backtrack or lookahead to see whether delay affects convergence and/or economy, violating local computation (12).

The principles (16), (18) and (19) constitute the core of our Minimal Computation Theory.¹⁸ The motivation is primarily conceptual, with the goal of reducing computational complexity, purporting the Optimality Hypothesis. Let's put them to test.

¹⁸I should note that the principles outlined here are largely compatible with the conceptual suggestions of Chomsky (1995: Chapter 4) and Collins's work on local economy (1997). There are important differences, however, particularly the treatment of interpretability (e.g. section 2.2.2) and my proposal of Feature Movement in overt syntax. Both of them, on my view, are directly related to computational analysis. I will highlight my proposals that differ from their work.

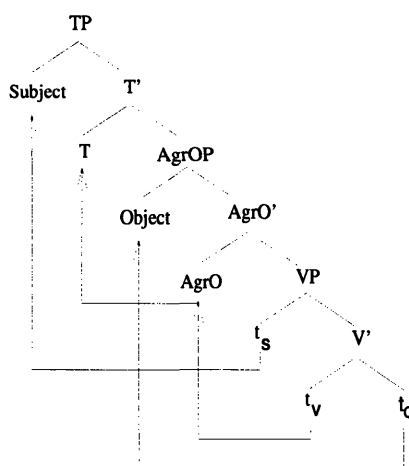
Chapter 3

Expletives and Economy

Chapter 2 laid out the conceptual foundations of the Minimal Computation Theory. The paradigm case for our empirical study is the expletive construction.

I assume the general machineries of Minimalist Program as in Chomsky and Lasnik (1993) and Chomsky (1993, 1995). Assume the simple transitive VP structure (20) in the Chomsky (1993: MPLT henceforth) with the functional categories (AGR and T) extended at LF:¹

(20)



Simple transitive VP at LF.

¹In fact, these functional categories are created as the derivation proceeds. Later developments, for example, the multiple-specifier model as in Chomsky (1995) and Ura (1996) are compatible with the proposals and analyses here.

For the simple transitive clause, V raises to AGRo then to T. For English, the Extended Projection Principle (EPP) (Chomsky 1981, 1995) requires [Spec,TP] to be structurally filled. Subject raises to [Spec,TP] to check off nominative Case and EPP. Further V-movement is possible and necessary for Multiple Subject Constructions (Chomsky 1995, Bobaljik and Jonas 1993, 1996), to which we return in Chapter 5. Since the nominative Case is checked with T raising, it is natural to suppose that infinite verbs can not check the nominative Case because of its degenerate (tenseless) inflection (Watanabe 1993) – consistent with traditional assumptions. At LF, accusative Case is checked in [Spec,AgrOP] with object raising.² Case checking is uniformly a Spec-Head relation. The Case Filter that every argument gets abstract Case holds at LF (Chomsky and Lasnik 1993).

The crucial issue here is that there should be no “optional” checking that involves distinct linguistic levels:³ Case is either checked in the overt syntax or at LF. This must be so because if optional operation enters into economy / convergence competition, it in general introduces computational intractability, as noted before. One cannot allow LF covert movements (even if they are “cheaper”) to compete in derivation. Therefore, Johnson and Lappin’s concerns about combinatorial explosion (1996: section 3.1) don’t arise.⁴ All syntactic operations are deterministic.

3.1 Expletives and the MPLT Analysis

Expletive Construction has been a focal point for some intensive studies in the MP framework (Bobaljik and Jonas 1993, 1996, Chomsky 1993, 1995, Frampton 1995, Groat 1995, Lasnik 1995, Collins 1997, Jang 1997, to name a few). The interaction of economy principles is directly relevant to complexity considerations. It is therefore

²But see Johnson (1991) and Koizumi (1995), among others, who propose that object moves in overt syntax. See section 4.1.

³See Chapter 2, section 3 and footnote 12.

⁴Specifically, they assume that the computation compares derivations with optional LF movements for subject, object, and T, etc. Therefore, a simple sentence “John likes Mary” will have a handful of competing derivations to co-exist, which, as they argue correctly, leads to intractability for examples like (11).

the central case of study for our analysis. Let's begin by reviewing Chomsky's analysis in MPLT.

Consider the expletive constructions⁵ in (21):

- (21) (a) there is [α a strange man] in the garden.
(b) \star there seems to [α a strange man] that it is raining outside.

I summarize the MPLT assumptions as follows:⁶

- (22) (a) α , not being in a proper position to check Case, must do so at LF, by raising to adjoin the LF affix *there*;

⁵It is clear that expletive construction is limited to a restricted set of verbs, perhaps, the unaccusative class and *be* (Belletti 1988):

- i. a. there arrived a man.
b. there was a unicorn in the garden.
c. \star there John loved Mary.

Some other restrictions are shown in the so-called Passive Expletive Constructions (PEC) for transitive verbs:

- ii. a. there has been a book put on the table.
b. \star there has been put a book on the table.

and the so-called Transitive Expletive Constructions (TEC) in Icelandic (Bobaljik and Jonas 1996), marginally in English:

- iii. a. There have many Christmas trolls eaten pudding. (Icelandic)
b. ? There shot the target a bullet.

I will discuss PEC and TEC in Chapter 5. For present purposes, we only consider the case of unaccusatives.

Also note that the pleonastic *it* exhibits quite different distributional properties from *there*.

- iv. a. \star it arrived a man.
b. \star it has been a book put on the table.

and

- v. a. \star there seems that Bill likes French fries.
b. it seems that Bill likes French fries.

Drawing from data on agreement, McCloskey (1991) notes that *it* does not involve expletive-associate linking. The contrast between *there* and *it* can be captured if we adopt the proposal that *it* is an argument (e.g. Koster 1986: p262). Throughout this paper, I will use the term "expletive" to denote the "pure" expletive *there* which has no Case or ϕ features nor does it require a θ -role. We will briefly return to the case of *it* in footnote 11 of Chapter 4.

⁶The subsequent analyses in Chomsky (1994, 1995) are roughly the same. Yet another one suggested by Chomsky (personal communication) is presented and adopted (with revision) in Chapter 4.

- (b) the expletive *there* must be coupled with an associate NP (α) that bears a theta role, for without it *there* would be a free-standing LF object that receives no interpretation at LF, violating FI;
- (c) *there* has no Case or ϕ features, but has a D-feature (1995: p364) thus is sufficient to check off the EPP feature.

For the MPLT analysis, α in (21a) raises at LF to *there*, checking off Case and forming the expletive-associate complex. In (21b), since α already has its Case feature checked in the complement position of *to*, it needs not (thus can not) raise at LF. Furthermore, since *there* checks off the matrix EPP, the derivation is legitimate but uninterpretable, as *there* is a free standing LF object. Thus (21b) is ruled out by FI at LF.

Before we go on, let's look at some data on locative inversion and its correlation to expletive constructions. This builds up the background for later discussions. (23) shows the unaccusative restriction shared by both:

- (23) (a) there is a man in the room.
in the room is a man.
- (b) ★ through the wedding band shot a marksman. (Bresnan 1994: 84)
★ there shot a marksman.
- (c) through the wedding band shot a bullet. (Bresnan 1994: 84)
there shot a bullet.

(24) shows that the complement/adjunct restriction in locative inversion (Bresnan 1994) carries over to expletive constructions:

- (24) (a) ★ Into the hole excreted the rabbit.
Into the hole jumped the rabbit.
(Bresnan 1994)
- (b) ★ There excreted the rabbit.
there jumped the rabbit.

(25) shows the Definiteness Effects (DE) for the associate subject, (controversially)

characteristic of expletive existentials as noted by Milsark (1974), Safir (1982), Belletti (1988) and others:⁷

- (25) (a) there came three armed aliens.
★ there came the three armed aliens.
- (b) From Mars came three armed aliens.
★ From Mars came the three armed aliens.

This correlation, I believe, is too strong to be coincidental. Indeed, Bresnan (1994) argues that the PP is the subject in locative inversion, hence patterns with *there*, which is also in the subject position [Spec,TP]. Levin and Rappoport (1995: p219-220) also conclude that *there*-existentials strongly correlate with locative inversion. Therefore, I will assume that this correlation is an *identity* relation, hence calls for a unified analysis. If this is true, for locative inversion, the MPLT analysis would say that the locative PP occupies the [Spec,TP] to satisfy the EPP feature, and the inverted subject raises at LF to check off nominative Case – essentially the analysis in Collins (1997) and Jang (1997). I will return to this in section 5.1.1.

⁷A word on the Definiteness Effects. Attempts have been made to account for this as a *syntactic* restriction. For example, Chomsky (1995: p342) suggests that T attracts only [N] feature but not [D] feature, the latter is assumed to be associated with specificity. But as Chomsky himself notes, there are some potential problems with this. For instance, in some constructions, violation of the Definiteness Effect seems to be quite acceptable:

- i. a. Into the lecture hall came the professor.
There came the professor.
b. There came the knight who has slain the dragon.
Into the village came the knight who has slain the dragon

In fact, post-verbal subject displacement in some other languages has no such effects:

- ii. ha mangiato Gianni (Italian)
Gianni ate.

Here Chomsky's N-to-T attraction analysis will not work, because *Gianni* is definite. Based on these observations, I think it might be appropriate to regard the Definiteness Effect as an interpretation constraint, following Milsark (1977) and Chomsky (1977). Note that in the unacceptable *There is John in the game*, morphological features (Case, EPP, and so on) are properly checked. See Chomsky (1995: p350 and fn 42, 44) for some suggestions. I would like to thank Howard Lasnik, Julie Legate, and Carolyn Smallwood for this discussion.

3.2 Objections to the MPLT analysis

There are a number of conceptual and empirical problems with the MPLT analysis. These problems are not particular to the MPLT analysis; in fact, they are, in one form or another, shared by all the analyses reviewed here. I refer to the MPLT analysis only because it is most clearly articulated. The problems are mutually related, but let me discuss them in turn.

3.2.1 Objection # 1: Optionality

Case

The MPLT analysis introduces optionality to nominative Case checking. Consider the following pair:

- (26) (a) a man is [_{sc} t in the room].
(b) there is [_{sc} a man in the room].

In (26a), *a man* raises out of the small clause to [Spec,TP] to check Case in overt syntax. Since the nominative Case checking can take place at LF, it is delayed to LF in (26b). The difference is in the EPP. The EPP feature is checked by *a man* in (26a) and *there* in (26b), respectively. So there is no reason to raise *a man* overtly. However, this optionality in Case checking makes it difficult to rule out (27):

- (27) ★ there [_{VP} a man likes Mary]

If *there* suffices to check off the EPP feature, then *a man* has no reason to move out of the VP internal shell, given that it is possible at LF (perhaps also cheaper). At LF, it directly raises to [Spec,TP] to check off Case, perhaps forming an expletive-associate pair as well. Thus, the derivation should converge, contrary to the fact. See also Lasnik (1995: p625-626) for similar arguments.

Conceptually, as we argued earlier, optionality induces complexity: the computation cannot decide locally which option to take, because it cannot in general predict the consequences of such options. In an optimal theory, optionality should be banned systematically.

EPP

Consider next the notion of EPP which requires [Spec,TP] to be filled. Let's assume, following Chomsky (1995), that *there* intrinsically lacks Case or ϕ features (but see Lasnik (1995) for a different view). Consider the other assumption in the MPLT analysis (22c), that *there* alone suffices to check off the EPP. A number of difficulties arise.

First, the insertion of *there* into [Spec,TP] to check off EPP runs against the general assumption that typically, feature checking is satisfied by Movement⁸. For instance, Chomsky's Chain Condition (1986) requires that the head of an A-chain must be in a Case position. Merge, on the other hand, is reserved for theta relations, essentially a base property, following Hale and Keyser's configurational approach to argument structures (1993). This Merge/Move distinction is pointed out by Chomsky (1995: section 4.6).

Second, there seems to be some inconsistency in the MPLT analysis from the assumptions in (22). Consider (28):

(28) I expected [there to be a book on the shelf]. (=61c, Chomsky 1995: Chapter4)

The derivation of (28) goes as follows. When the TP in (29) is formed, the EPP feature of TP needs to be checked off:

(29) [_{TP} to be a book on the shelf]

Overtly raising *a book* is a possibility. However, insertion of *there* is also available, given that *there* alone checks off the EPP feature (22c). Since merge is cheaper than overt movement, and Case checking can be delayed to LF per Procrastinate, (30) is formed:

(30) [_{TP} there [_{T'} to be a book on the shelf]

Assume that the ECM Case is checked in [Spec,AgrOP] at LF⁹ (Chomsky and Lasnik 1993). Thus, at LF, we have the following structure:

⁸That is, *if* the EPP is indeed a feature. I will argue otherwise in Chapter 4.

⁹An assumption that is not entirely unproblematic, to which I return in Chapter 4. It does not affect the argumentation here, however.

(31) I expected [_{AgrOP} [_α a book] [_{TP} there to be t_α on the shelf]]

In recent editions of Minimalist Program, Chomsky (1995) argued that feature checking should be conceived as attraction, which directly incorporates economy principles such as the Minimal Link Condition that the closest candidate for feature checking gets attracted. In this case, it is *a book* (or its features), which is the only candidate. Note that *there* becomes a free-standing LF object, because *a book* directly goes into [_{Spec,AgrOP}] to check off accusative Case, skipping the Spec of the embedded TP hence *there*. It is unclear how *a book* forms the coupling relation with *there* as a complex.

One might suppose that *there* has a [-N] feature that attracts the N-feature of its associate (Chomsky 1995, Frampton 1995, and Jang 1997), and therefore the expletive-associate relation is mandatory. This is also problematic, for it incorrectly admits:

(32) * [there a man]

In fact, as observed by Chomsky (1995: p362) and Jang (1997: p61), the expletive is always situated in [_{Spec,TP}] and it is in [_{Spec,TP}] where the the expletive-associate pairing takes place.¹⁰ This observation makes suspect the idea that it is the expletive that intrinsically attracts the associate.¹¹ On the other hand, if the unification of locative inversion with expletive construction is possible, it is hard to see why and how the locative PP should attract the post-verbal NP, because PP is interpretable at LF and thus requires no associate.

The essence of the MPLT analysis seems correct: as a result of FI, expletives must be coupled with an associate, otherwise it becomes a free-standing uninterpretable LF

¹⁰Lasnik (1995) makes the similar point, but for different reasons. He notes the contrast between the following minimal pair (his 14):

- i. a. I want there to be someone here at 6:00.
- b. * I want there someone here at 6:00.

For Lasnik, the contrast is due to the fact that *someone* in (i.a) bears partitive Case from *be*, thus eligible to adjoin to *there*. It is possible to explain this contrast in terms of T without resorting to partitive Case; see section 4.3 and (67d).

¹¹In Chapter 5, I will propose that it is T that is doing the attraction. The expletive-associate pairing is indeed formed in [_{Spec,TP}], but for independent reasons.

object. However, by assuming that *there* alone is sufficient for EPP checking (22c), and that nominative Case checking can take place at LF, this expletive-associate is broken as shown in (28). The MPLT intuition has to be materialized differently.

Agreement

This objection is speculative and somehow theory-internal but I believe still worth considering. It also lays out some background for the discussion in Chapter 4.

There are reasons to believe that nominative Case and agreement are two sides of the same coin, regardless of the position of the subject.

- (33) (a) a man is / (★ are) in the room.
 there is / (★ are) a man in the room.
- (b) È arrivato Gianni.
 is arrived Gianni.
 “Gianni has arrived”.
- (c) Ha telefonato sua moglie.
 has telephoned your wife.
 “Your wife has telephoned”

Consider also following examples in Icelandic (Frampton 1995):

- (34) (a) Bókin var gefin okkur.
 book-the(Nom) was(3rg) given us(Dat).
- (b) Okkur var gefin bókin.
 us (Dat) was(3rg) given book-the(Nom).

In (34a), the indirect object receives inherent Case directly and the direct object moves to [Spec,AgrSP] for nominative Case, correlating with agreement. In (34b), the indirect object again receives inherent case, but moves to [Spec,TP] to check off the EPP feature. Presumably, the direct object checks off nominative Case at LF (under the MPLT analysis) – and agreement pairs with it. Hence, as an observation,

agreement seems to go hand in hand with subject that checks nominative Case (cf. Safir (1985: p193-203)).

According to the MPLT analysis, displaced subject (in expletive constructions) checks Case at LF; however, agreement is manifested at PF. This is not really a problem, since MPLT assumes a strong lexicalist theory of morphology. That is, the phonetic features of lexical items are present in the numeration and do enter into the syntactic derivation. Consider (35):

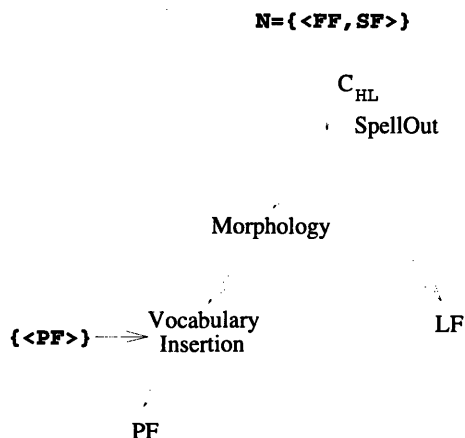
- (35) (a) There is a man in the room.
(b) ★ There are a man in the room.

The agreement features are “base-generated” in *is/are*, given in the numeration. There is no need for the subject to affect the agreement features: a mismatch simply cancels the derivation of (35b). To this extent, the MPLT analysis is perfectly consistent.

The LF Case and PF agreement correlation only becomes problematic with respect to the Distributed Morphology ¹² (Halle and Marantz 1993, 1994). A key feature of Distributed Morphology is Late Insertion: phonological features are supplied *after syntax*, by insertion of Vocabulary Items into the terminal nodes. Late Insertion adds phonological features to terminal nodes, but it does not add to the semantic/syntactic features making up the terminal nodes. This contrasts with the lexicalist version of the Minimalist Program, for which the numeration *N* consists of a set of lexical items, each of which is a bundle of Phonetic, Semantic, and Formal features. If there are independent reasons to adopt Late Insertion, then the only features active during the (pre-Morphology) syntax are Formal and Semantic ones. Phonetic Features are inserted, following instructions from syntax and morphology to the perceptual/articulatory system. Combining (1) and Halle and Marantz (1994: figure 2), the organization of grammar is like this (see also Sauerland 1996):

¹²I would like to thank Martha McGinnis for a discussion of the materials presented here.

(36)



Minimalist Program meets Distributed Morphology

If this conception of grammar is correct, it is impossible for LF at the Conceptual-Intentional interface to give instructions to the morphological or phonological components, since the computation has already branched out at Spell-Out. Hence, if overt PF agreement correlates with nominative Case checking, then *before* Spell-Out, the C_{HL} must have provided Vocabulary Insertion with instructions to add Phonetic Features to the verb nodes (cf. Chomsky (1995: p289)). Hence, nominative Case checking must take place before Spell-Out.

One might argue that agreement is a result of ϕ -feature checking. Therefore, agreement is manifested *independent* of Case checking. In the model of grammar (36), the ϕ -features of subject move to Agr or T, before Spell-Out, hence we obtain overt agreement. This move is conceptually undesirable. Recall that in the Minimalist framework, the “motivation” for syntactic operations is to check off (eliminate) uninterpretable morphological features. Agreement, on the other hand, has to do with relations among *semantic* features, which, crucially, do not delete since they are interpretable. In this sense, agreement should not be a driving force for operations. Instead, it is perhaps a by-product, namely, a featural relation that just “happens” under certain structural configurations, say, as the Spec-Head relation between the ϕ -features of the Case-checking head and the checker (subject). If so, the apparent

divorce of agreement (manifested in overt syntax, before Spell-Out) and nominative Case checking (at LF, after Spell-Out) becomes a mystery.

In summary, by assuming that nominative Case checking can take place at LF, optionality is introduced. Of course, one can always assume that nominative Case checking at LF is only available when an expletive *there* in the numeration. In this sense, the system avoids systematic optionality by using a computational trick: look at the numeration, if an expletive “there” is found, then nominative Case checking is delayed at LF. Besides the stipulative flavor, it is untenable. Given that *there* lacks intrinsic Case features (Chomsky 1995), why would it affect Case checking at all? Other improprieties remain, see (28) and (32) and the discussions there.

3.2.2 Objection # 2: Complexity

The second objection is both conceptual and empirical in nature. To recapitulate briefly, I have argued in Chapter 2 that upon availability, the theory only generates convergent thus interpretable derivations. Furthermore, the derivation must be carried out efficiently, purporting the Optimality Hypothesis. A consequence of this is that we must reject a theory for which uninterpretable derivations can be generated, albeit locally. In the optimal theory, local economy must ensure convergence.

I will review two analyses that involve global computations: the MPLT-type that is based LF Case checking and the partitive Case of Belletti (1988), Lasnik (1995), and Johnson and Lappin (1996).

MPLT

Consider the following minimal pairs (Chomsky 1994, 1995):

- (37) (a) There seems to be a man in the room.
(b) ★ There seems a man to be in the room.
- (38) (a) ★ Mary believes to be a man in the room.
(b) Mary believes a man to be in the room.

The MPLT analysis is as follows. First, the small clause [*a man in the room*] is merged with *to be* to form:

(39) [_{T'} *to be* [_{SC} *a man in the room*]]

Now the EPP feature of T is to be satisfied. Two options: (a) merging *there* into [_{Spec,TP}] from the numeration, or (b) overtly raising *a man*. Overt raising violates Procrastinate, since it is not required for convergence, provided that it can raise at LF to check off Case. Hence *there* is merged to construct (40). Hence (37b) is banned.

(40) [_{TP} *there to be* [_{SC} *a man in the room*]]

The structure in (40) is subsequently merged with *seems*:

(41) [_{T'} *seems* [_{TP} *there to be* [_{SC} *a man in the room*]]]

Here the only operation available is the raising of *there* to [_{Spec,TP}], given the Minimal Link Condition, for it is closer to T than *a man*. This yields (37a). At LF, *a man* raises to check off nominative Case, adjoining to “there” to form the expletive-associate pair.

The paradigm in (38) seems somewhat unexpected. Consider Chomsky’s analysis (1995: p347). When the embedded TP is formed, the computation again has two options: either to raise *a man* or to insert *Mary* from the numeration. For the same reason as before, *Mary* is inserted:¹³

(42) [_{TP} *Mary to be* [_{SC} *a man in the room*]]

Indeed, this decision is made locally. Although it is cheaper, it is nonetheless wrong: merging *Mary* leads to an uninterpretable structure:

(43) [_{TP} *Mary_t believes* [_{TP} *t to be* [_{SC} *a man in the room*]]]

Now *Mary* moves from the embedded to the matrix subject position but still lacks a theta role, crashing at LF. Chomsky then assumes that economy is overridden

¹³But see footnote 6.

by convergence,¹⁴ i.e. economy only chooses among convergent derivations. Hence, economy conditions do not apply, and the computation will pursue the other option: raising *a man* to fill the embedded [Spec,TP]:

(44) [TP a man to be [SC t in the room]]

The rest of the derivation is obvious.

By assuming that economy is overridden by convergence (Procrastinate), we are drifting away from the Optimality Hypothesis: How does the computation “know”, at an intermediate stage of the derivation, whether a cheaper operation would eventually lead to divergence, or it should be overridden in favor of some more expensive but convergent derivations? In general, the computation cannot make this decision locally, as noted earlier in Chapter 2. In the optimal theory, economy should not be overridden by (global) convergence; instead, local economy should ensure global convergence: if economy conditions are obeyed, nothing but convergent derivations will be generated and the computation does not have to backtrack to pursue alternative operations. Departure from this introduces complexity and we are back to global economy again. This is obviously a strong claim than needs to be substantiated. I will try to do so in Chapter 4.

Let’s now briefly review two other analyses related to MPLT. Collins (1997) also notes the globality of Procrastinate in the MPLT analysis. He proposes an additional economy principle (p123):

(45) Chain Formation Principle 4: If there are two operations OP_1 and OP_2 applicable at a set of representations Σ (both satisfying Last Resort and Minimality), then choose the operation that extends an incomplete chain.

The basic idea is that although movement can check off the EPP feature of T, the primary motivation to form a chain is to check the Case assigning feature of T.¹⁵ Principle (45) is indeed a local economy principle, but I don’t see how it accounts for the paradigm in (38). Consider again the intermediate stage of the derivation:

¹⁴Note that the term “convergence” here includes “legitimacy”, all defined in (3).

¹⁵See Collins (1997) for details. Here I put aside his considerations for the [+wh] and Q features of C in chain formation, irrelevant to the discussion here.

(46) [T' to be [SC a man in the room]]

Collins notes (p124-125) that

At this point, raising of *a man* (Copy + Merge) would satisfy Last Resort, so it is allowed. Other economical options would include inserting *Mary* (Copy from the lexicon, plus Merge). However, *this operation would eventually lead to an uninterpretable structure, so there is no need to pursue it* (my emphasis: CDY).

There are some problems with (45) and the statement quoted above. When the TP in (46) is formed, there are two operations available to check off the EPP feature: inserting *Mary* or moving *a man*. It appears impossible to predict, locally, which of these two operations completes a chain. Of course, raising **eventually** completes a chain with *a man* in [Spec,AgrOP] at LF, but there is no way to know this at an intermediate stage of the derivation. Since Collins's analysis allows the insertion of *Mary* as a possibility, it is unclear how the computation "knows" inserting *Mary* eventually leads to an uninterpretable structure (as emphasized in the quote above). In my view, the computational system cannot make this decision locally, unless it looks ahead: therefore (38a) crashes and the computation has to go back to pursue (38b). As noted in section 2.2.2, when uninterpretable hence divergent derivations are allowed, the Optimality Hypothesis is violated.

Consider next Geoff Poole's analysis (1995). He proposes the following condition:

(47) Total Checking Principle (TCP): The most economical operation is one where an element moves to, or is inserted into, a position where all of the formal features it bears enters into a checking position.

This is also a local principle and indeed accounts for (37). When the embedded TP is formed, the computation has to choose between inserting *there* and raising *a man* to satisfy the embedded EPP feature. The latter violates the TCP, because its Case feature is not checked. The former does not, since the pure expletive *there* does not

even have Case or ϕ features but only has a D feature,¹⁶ and this D feature (its only feature) is checked in the process. Thus, *there* is inserted.

The problem with Poole's TCP, in my understanding, is that it does not account for the pair (38) in strict local economy. In the derivation, the computation again chooses between merging *Mary* and raising *a man*. However, these two options violate the TCP **to the same degree**: they all check off the EPP feature of the embedded TP and none of them checks off the Case feature of itself – unless global computation is invoked, which involves look-ahead: raising *a man* eventually checks off accusative Case at LF. No local decision can be made since the two operations have **the same cost** in the sense of the TCP and cannot be distinguished. Poole's analysis thus does not go through.

The Partitive Case Analysis

The second class of analysis I will consider is Johnson and Lappin's approach (1996) with a modification to partitive Case analysis of Belletti (1988) and Lasnik (1995). They assume that *be* lexically checks the partitive case of its NP complement in situ. In (37b), *a man* cannot receive the partitive Case since it has moved to [Spec,T] (presumably to check the EPP). It cannot check Case from matrix T either, because *seem*, unlike ECM verbs, does not check accusative Case for the embedded subject. Therefore, (37b) is a Case violation.

None of these authors discussed the paradigm (38) in details. Shalom Lappin (personal communication) did offer the following analysis. For (38a), he suggested that if *Mary* is inserted in the embedded subject position and then raises to the matrix, it could be ruled out by the Theta Criterion (FI) at LF: that *Mary* does not have a theta role in the chain thus formed.¹⁷ However, as argued earlier in Chapter 2, when FI is referenced only at LF (post Spell-Out), convergent but uninterpretable deriva-

¹⁶Poole's assumption, also Chomsky (1995), Frampton (1995) and others, which I reject in Chapter 4.

¹⁷Note that (38a) is not an EPP violation, for the EPP feature of the TP = [*to be a man in the room*] has been satisfied by inserting *Mary*, which then raises to the matrix subject position. It checks off the EPP features of both TPs (embedded and matrix).

tions are allowed, violating the Optimality Hypothesis. Alternatively, he suggested that if *Mary* is directly inserted into the matrix subject position, (38a) would be an EPP violation, since the embedded subject position is unfilled. This is again problematic, since it requires “postponement” of feature checking. Specifically, according to Lappin, when the embedded TP = [*to be a man in the room*] is formed, its EPP feature is not checked right away. Instead, the TP is merged with the matrix verb. This is dubious because massive computational complexity is introduced: how does the computation predict the consequences of postponing feature checking in general? In this case, it leads to the undesired divergence, while a convergent derivation does exist. In general, checking cannot be optional or delayed. Instead, something like the ASAP (19) with a greedy character should be in place.

For (38b), Lappin states that the post-copular NP complement of *be* moves to [Spec,TP] only if it does not receive partitive Case in-situ. Therefore, the fact that *a man* is in the embedded [Spec,TP] of (38b) entails that it receives accusative Case from the matrix verb *believes*. This violates local computation. When the TP = [*to be a man in the room*] is formed but **before** it is merged with the matrix verb, the computation faces a dilemma: either to leave *a man* in-situ to check partitive Case, or to raise it to [Spec,TP] hoping that it will check (accusative) Case later. In either case, the decision cannot be made locally. Keeping *a man* in-situ for partitive Case leaves the matrix accusative Case unchecked, crashing at LF – a non-local stage. Alternatively, raising *a man* would require the computation to “foresee” that an ECM verb is available, because the subject position of an infinitival is in general a non-Case position (can not raise to [Spec,AgrOP] to check accusative Case of ECM verbs). More fundamentally, if one assumes that movement is always Last Resort (don’t move if don’t have to), *a man* **will** stay in-situ, since the partitive Case checking is available: (38b) cannot be derived. In short, the partitive analysis does not seem to be a local (or coherent) solution to the paradigm in (37) and (38), and has roughly the same kind of complexity problems as in MPLT, by directly invoking LF conditions or lookahead.

To conclude this chapter, I have shown that the MPLT and the partitive Case analyses involve global complexity and/or have some inconsistencies. Recall that just

local computation is not good enough. Local economy must ensure global economy, and on top of all, it must satisfy the interface conditions and derive only convergent structures. In what follows, I will propose a solution to resolve these problems.

Chapter 4

T

To account for the expletive paradigms and to avoid the problems noted in Chapter 3, I will propose a generalized theory of feature movement in the Attract-F framework. Specifically, only features required for local checking get attracted. When PF features are not required, they are not attracted: only the relevant formal/semantic features are, even in “overt syntax”, i.e. pre-SpellOut, as briefly mentioned in section 2.4. Then I will propose to reconsider the nature of the Extended Projection Principle (EPP) and show that the traditional notion of EPP (Chomsky 1981) should be eliminated in favor of the T (Case) system and its peripheral PF properties. Many desirable consequences follow, shown in Chapter 5, including a unification of languages and constructions that involve displaced and null subjects. Let’s start with some background observations.

The objections to the MPLT analysis can be summarized as follows:

- (48) (a) the optionality of Case checking at LF, and the computational complexity it induces
- (b) the broken pairing between the expletive *there* and the associate *a book* in examples such as (28)
- (c) tentatively, the divorce between agreement and nominative Case checking, in light of the Distributed Morphology

Our problem is the following. On the one hand, we wish to eliminate the option-

ality in nominative Case checking. On the other, we have to establish the expletive-associate pairing at some point of the derivation. Note that the latter occurs in [Spec,TP], as an observation. However, there is no compelling reason to assume that the expletive attracts the associate; see (32). If *there* suffices to check off the EPP, then the associate NP would “skip” the expletive and get attracted to Case positions directly; see (28). The solution, I think, lies in some revision to the current Minimalist system.

4.1 Minimal Feature Attraction

I will first advocate a natural extension of the checking theory in the Attract-F framework. Specifically, I propose the following:

- (49) Minimal Feature Attraction (MFA). An operation for checking off feature F attracts just enough the features that suffice to check off F .

Of course, (49) is subject to Last Resort (16) and other general economy principles. The MLC dictates that only the closest feature(s) be attracted. Hence, the MFA can be viewed as part of the Minimality condition. The MFA is fundamentally different from claims such as “attract just enough features for **convergence**”, as noted in section 2.4. Reference to convergence is in general global and must be rejected in a local theory. The MFA simply attracts the minimal set of closest features for checking F , upon availability.

The idea of MFA is not new. Its theoretical antecedent can be traced back to (at least) the expletive-replacement analysis in Chomsky (1986). It has been long noted that in existential constructions, associate NP creates binding possibilities:

- (50) There are always linguists criticizing each other’s work.

Lasnik and Saito (1991) noted that direct objects of transitive objects can bind into adjuncts:

- (51) I saw two men on each other’s birthdays.

This suggest that somehow the NP becomes a binder. However, literal replacement of expletive by associate doesn't seem to be an adequate solution. Chomsky (1991) notes a scoping difference:

(52) There aren't many linguistics students here.

If *many linguistics students* substitutes for *there*, it would have wide scope with respect to negation. The fact is, it has only narrow scope. Hence, it has been suggested that at LF, only features that are relevant for checking raise. In (52), the category of the associate stays in-situ, only the relevant features raise to check Case, including the semantic features that cause the binding effects. Chomsky (1995: p264-266) and Lasnik (1996) also propose that only necessary features are attracted in checking. Jang (1997) pursued this idea in some details and gave an account account for the Definiteness Effect, scope, and negative polarity licensing in expletive constructions.

If this is correct, the immediate question is: Why should this simple and natural principle apply only at LF? This is curious since movement is conceived as mechanical feature attraction. Of particular interest are PF features. Conceptually, they should never move, unless forced by PF conditions, heard in the output. This condition should not be stated disjunctively for different linguistic levels (LF and overt syntax). Therefore I assume that Minimal Feature Attraction applies **uniformly** throughout the computation. It is possible for formal and semantic features to move alone without pied-piping phonetic features, in overt syntax.

Before returning to expletive constructions, I should mention that if the MFA is true, that feature attraction suffices for checking, many theoretical issues will have to be reformulated. I will only briefly consider that of Case checking here. It has been suggested that in English, accusative Case checking occurs at LF (Lasnik and Saito 1991, Chomsky and Lasnik 1993):

- (53) (a) The DA proved [the defendants to be guilty] during each other's trials.
(b) ★The DA proved [that the defendants were guilty] during each other's trials.

Lasnik and Saito note that the subject of the infinitival complements in (53a) can bind the anaphor in the matrix adjunct, but not when the complement is a finite clause (53b). This suggests that the subject of infinitival moves to the matrix AgrOP at LF to have its Case checked. However, Jang (1997: p28), citing personal communication with Noam Chomsky, notes that the binding effects in (53) are not conclusive in favor of the LF Case checking:

- (54) (a) Damaging evidence about the men came to light during each other's trials.
(b) The DA presented damaging evidence about the men during each other's trials.

Chomsky notes that the judgment of (54) is on the same par with (53a) to be of the same marginal acceptability and that this is not relevant to object raising at LF, contra Lasnik and Saito (1991). Instead, the lack of binding effects is due to the fact that the c-command relation is weakened for phrases like *each other's XP*.

Under the MFA, a new interpretation is possible: the formal features of object raise to [Spec,AgrOP], perhaps carrying along relevant semantic features, in overt syntax. Here, *overt syntax* has nothing to do with overtness; it is simply a technical term that refers to computations **before** Spell-Out. By conceptual hypothesis, PF features are attracted only when required by interface conditions. A feature is Strong/Weak if it does/doesn't require a PF category. In English, nominative Case is strong and therefore requires an overt category; accusative Case checking is weak, thus pure feature movement suffices and no visible object movement is possible or necessary. Since the formal/semantic features of the object are situated in [Spec,AgrOP], binding and scoping effects follow.¹ In this sense, word order literally reduces to PF feature strength. All Case checking involves feature raising before Spell-Out. For SOV languages, both nominative and accusative Cases are strong. For VSO, both are weak (see Stabler 1996 for some similar ideas). In any case, the structure at LF should be

¹cf. Koizumi (1993) and Lasnik (1995a), who argue that the entire object constituent raises in syntax. It is possible to restate their proposals here, since all relevant grammatical features raise to [Spec,AgrOP] in syntax, hence the binding effects and so on. The PF features of the object remain in-situ. More recently, Lasnik (1995a, 1996) has argued for a feature-raising analysis as well.

identical for all languages, with formal and semantic features situated and checked in their respective positions.

The MFA perhaps generalizes to Wh movement as well (in fact, any movement), similar in spirit to Watanabe (1991). Suppose that Wh movement is driven by the [-wh] feature of C, perhaps successively as in Spanish (Torrego 1984), Irish (McClosky 1989), Ewe (Collins 1993), and many other languages. On the feature-movement view, parametric variations in Wh movement for English/Japanese reduce to feature strength, a PF condition. For English, C has a *strong* [-wh] feature, which triggers overt wh-movement; for Japanese, C has a *weak* [-wh] feature, which requires only movement of Wh features but no overt category. All languages, with pronounced Wh movement or not, all have overt [wh] feature movement and therefore have identical LF representations. Again, language variations reduce to parametric PF interface conditions, natural to the Minimalist intuition.²

Back to expletives. The MFA in overt syntax is often obscured by the strong Case feature, thus sometimes “invisible”. A strong feature requires some overt category for checking. One such requirement is the EPP of English, according to which [Spec,TP] must be overtly filled. Consider the simple transitive:

(55) [_{VP} John likes Mary]

As discussed and desired earlier, nominative Case checking should occur before Spell-Out. For (55), the derivation raises *John* to [Spec,TP] to check off the EPP and nominative Case simultaneously:

(56) [_{TP} John [t likes Mary]]
Case/EPP

However, there is still a crucial piece missing in our puzzle. What is the driving force for the feature movement of the associate in expletive constructions? Here the features cannot move as free-riders since the subject category remains in-situ and

²Note that the MFA is in harmony with the idea of Late Insertion in the Distributed Morphology (36). In both languages, the PF parameters give instructions to the phonological component: for English, add the Phonetic features to the terminal nodes that contain [+Wh] features, and for Japanese, don't.

thus does not pied-pipe. Moreover, note that in the MPLT analysis, the motivation cannot be EPP: the EPP is satisfied by the expletive *there*, and has nothing to do with the associate. Consider the problematic example (28), repeated here as (57):

(57) I expected there to be a book on the shelf.

As noted earlier, the N-to-D raising of the associate to the expletive is untenable. Therefore, if *a man* directly raises to [Spec,AgrOP] to check Case, it will “skip” *there*, failing to form the expletive-associate pair thus violating FI. However, (57) is perfect. It seems that as soon as *there* enters in the derivation at a particular position, it must form a coupling relation with its associate, precisely **at that position**. In this case, the position is [Spec,TP], where the so-called EPP feature is checked.

4.2 The nature of T

“The *Extended Projection Principle* (EPP) states that [Spec,IP] is obligatory, perhaps as a morphological property of I or by virtue of the predication character of VP (Williams 1980, Rothstein 1983)” (Chomsky and Lasnik 1993, in Chomsky 1995: p55). Even since its proposal (Chomsky 1981), the nature of EPP has been controversial, particularly with respect to “pro-drop” languages, which allow null subjects, and some VSO languages, such as Irish, which **cannot** have pre-verbal subjects. Even for English, a wide range of complications arise. To be sure, the [Spec,IP] (“subject”) position is at least somewhat different from the VP complement position, which is always a θ -position. [Spec,IP] is only an A-position, a **potential** θ -position, depending on the lexical choice of the V. Both an expletive and an NP can appear in that position:

- (58) (a) There is a man in the room.
(b) A man is in the room.

Indeed, on surface, almost **anything**, not just arguments, can appear in [Spec,IP]:³

³See Hoekstra and Mulder 1990, Bresnan 1994, Collins 1997, Jang 1997, and others for arguments that both locative PP and quote appear in the subject position. In section 5.1.1, I will argue that this is not entirely accurate.

- (59) (a) Down the hill rolled John. (locative inversion)
 (b) “This apple is really sweet”, said Adam. (quotative inversion)
 (c) Sweet indeed was the apple Eve gave me. (adjective phrase)

It is thus tempting to conclude that the EPP is simply a requirement that [Spec,TP] must be overtly filled. It is in fact suggested by Chomsky (1995: p199) that the EPP can be reduced to a strong or weak NP feature. Following this line of argument, Jang (1997: Chapter 4) concludes, based on his analysis of locative inversion, that locative PP appears in [Spec,TP] for the simple “overtness” requirement of T. This accounts for the fact that, as noted above, almost anything can fill [Spec,TP]. For locative inversion, the element in [Spec,TP] is not restricted to DP, contra Collins’s assumption that it must be filled with a phrase with [D] feature (1997).⁴

This conclusion is conceptually appealing. The problem is that it still fails to capture the expletive-associate relation pointed out earlier. Specifically, consider the simple expletive construction *there is a man in the room*. When the embedded TP is formed:

- (60) [_{TP} is [a man in the room]]

According to this analysis, *there* suffices to check off the EPP feature (the “overtness” requirement) and is thus inserted. The derivation converges. However, as noted before (57), *a man* (or its formal features) will directly raise to check Case (in overt syntax under the uniform MFA framework, or at LF in other analyses), for there is no reason for it to form relation with the expletive. The D-to-N raising analysis does not work either; see (32).

I suggest a reconsideration of the EPP feature. The general intuition that English [Spec,TP] must be overtly filled seems to be correct. But we do wish to capture the observation that expletive-associate pairing must be established as soon as the expletive is introduced into the derivation, at [Spec,TP]. With the MFA, this relation is established via movement of features.

⁴Collins does suggest that, alternatively, [Spec,TP] could be filled with any category (p28), but offers no explanation.

(61) The Nature of T:

- (a) T universally attracts the closest available nominal features;⁵
- (b) that T is parametrically strong/weak.

(61) essentially decomposes the traditional notion of EPP into two modular interface conditions. The universal attraction of Case (and ϕ) features into [Spec,TP] (61a) is required by LF, assuming that the Case feature of T, if unchecked, causes the derivation to crash at LF. This is similar to a suggestion made by Chomsky (1995: p276) that the ϕ features of the associate raise to INFL instead of adjoining to the expletive. I will call the features attracted to T altogether *FF*, which certainly include the categorial and ϕ features. Although the attraction is universal, only tensed T **checks** (i.e. eliminates) the Case feature of the nominals, rendering them inert for further checking. (61b) says that the attraction is parametrically accompanied with the pronunciation of an overt category at PF.⁶ This captures this observation that the expletive-associate coupling takes place precisely at [Spec,TP] (INFL), because for languages like English, nominative Case is strong and requires an overt PF category. Crucially, it is T (INFL), but not the expletive that does the attraction, contra the assumption that T has a [-D] feature that is checked off by the expletive insertion, as in Chomsky (1995), Frampton (1995) and others.

In this view, the traditional EPP is not a feature anymore; rather, it is an instruction given to the Phonetic component of the grammar to insert some overt element in the position of [Spec,TP].⁷ The traditional notion is essentially eliminated, in favor

⁵It has been suggested, particularly in the later version of the Minimal Program (Chomsky 1995), that movement is triggered by categorial features of T (hence, not Case, a formal feature) to create checking relationships. As noted by Legate and Smallwood (1997), who examined object shift in Icelandic and subject-less clauses in Irish, the Case-based model is needed, for both conceptual and empirical reasons. Their argument hinges on the assumption in Chomsky (1995) that T has a categorial feature [-D], which induces the insertion of the expletive. If so, there would be no driving force by the attraction of the associate (at LF, presumably). See their discussion for details. Their challenge does not arise here, since I explicitly reject the idea that T has a [-D] feature. The insertion of expletive is triggered by strong T feature in English (61b). Hence, it is possible to maintain that Case checking is still mutual attraction of categorial features, namely, T universally attracts nominal features of DP/NP.

⁶I do not have a profound explanation why language is like this way, but only state it as a fact. The advantage of this proposal can be seen as we go along.

⁷I will still refer to this process as movement of PF features, only for convenience of discussion.

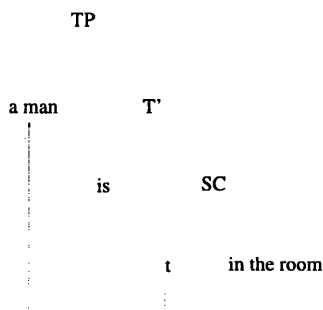
of the T system and its peripheral PF properties. In Chapter 5, I will consider some general consequences of (61), in association with the MFA (49). For now, let's go back to expletive constructions.

4.3 Expletives Revisited

The T system as stated a two-clause condition in (61), can be satisfied by either a single category altogether or separately by an overt phrase for (61b) together with the FF features from another category for (61a). Let's see how it interacts with the MFA to resolve some problems raised earlier.

In the simple (non-inverted) transitive case (*a man is in the room*), the PF features of *a man* raise from the embedded small clause to [Spec,TP], enforcing the overtness requirement in (61b). The FF of the subject are attracted to T to satisfy (61a):

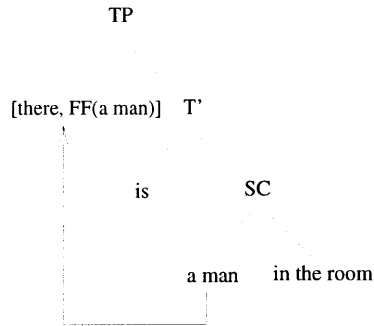
(62)



In the expletive construction, the two clauses of (61) are satisfied separately. The expletive is inserted to satisfy (61b), and the features of the subject (associate) raise to [Spec,TP] to satisfy (61a). Thus, a complex is formed at [Spec,TP] to establish the expletive-associate relation (63):⁸

⁸I use [*there*,FF(NP)] to denote that both *there* and FF(NP) are in the checking domain (Chomsky 1995) of T. I make no particular commitment to the configurational structure of the complex, though it is conceivable that a feature bundle as such is formed, with *there* providing the PF feature and FF(NP) the formal/semantic features.

(63)



One might ask, in the derivation of (63), what bans the overt raising of the subject, satisfying the two clauses of (61) altogether? This is directly related to the complexity considerations that in fact motivated this thesis. I will show that the MFA (49) and the two-clause EPP (61) cover paradigms problematic for other analyses, while keeping to minimal complexity and local computation. Before we start, I summarize my assumptions and proposals so far:

- (64) (a) Minimal Feature Attraction for (local) checking (49)
(b) nominal feature attraction to T is uniformly before SpellOut, and word order is determined by parametric PF conditions (61)
(c) *there* is a “pure” expletive, nothing but a PF feature bundle, only to satisfy (61b),
(d) Merge is costless, feature movement (FM) is cheaper than overt movement (OM), which moves more features:
 $0 = \text{cost}(\text{Merge}) < \text{cost}(\text{FM}) < \text{cost}(\text{OM}).$

Consider the following simple examples:

- (65) (a) There is a man in the room.
(b) A man is in the room.
(c) There arrived a man.

First consider (65a). The SC = [*a man in the room*] is merged with *is*, a finite verb. Now we must check off the EPP feature. The computation could raise *a man* overtly;

however, with *there* in the numeration, a cheaper operation exists. Recall that T checking consists of two parts: (a) an overt category for the strong feature, satisfied by *there* and (b) the Case feature for T, satisfied by feature raising. The cost of these two operations **accumulatively** is:

$$(66) \text{ cost(FM)} + \text{cost(Merge)} = \text{cost(FM)} + 0 = \text{cost(FM)} < \text{cost(OM)}$$

Therefore, the computation inserts *there* and raises *FF(a man)*, rather than raising *a man* overtly. Note that this choice is made **strictly locally**, without making references to other stages of the derivational, or trans-derivational comparisons; cf. (12). The derivation converges.

For (65b), since *there* is not present in the numeration, both *FF(a man)* and *PF(a man)* overtly raise to check off the Strong matrix T feature (Case) So the two clauses of (61) are satisfied altogether by *a man*. (65c) is similar to (65a): *a man* in the complement could have raised to [Spec,TP], if *there* were not present (which would derive *a man arrived*). Since *there* is in fact available, it is inserted into [Spec,TP], coupling with the movement of *FF(a man)* to check off the EPP feature. Note that for all the examples in (65), *FF(man)* raise into [Spec,TP]: altogether with the whole category in (b), and alone in (a,c). In both cases, overt agreement follows immediately: the ϕ -features of NP **are** in [Spec,TP], matching with those of T. This analysis is clearly compatible with the idea of Late Insertion in Distributed Morphology (cf. Chomsky 1995: p289 and footnote 61).

Consider then the following “troublemakers”:

- (67) (a) *There seems to a man that it is raining outside. (Chomsky 1993)
 (b) *There₁ seems there₂ is a man in the room. (Chomsky 1995)
 (c) I expected there to be a book on the shelf. (= (28)) (Chomsky 1995)
 (d) *I want there someone here at 6:00. (Lasnik 1995)

For (67a), the insertion of *there* into matrix [Spec,TP] is not sufficient, since the movement of *FF(a man)* is not available, having been checked off (deleted) in the complement position of P (*to*). Thus, (67a) is a Case violation for the matrix T. In

(67b), *there*₂ and *FF(a man)* check off the embedded strong T feature and *FF(man)* gets Case. Hence, *FF(a man)* can not raise again to *there*₁, for the same reasons as in (67a).⁹ (67c) is derived as follows. When the TP = [*to be a man in the room*] is formed, T requires both FF features and an overt category. Raising of *a book* is more costly than raising its formal features together with merging *there*. Hence “[*there, FF(a man)*] *to be a man in the room*” is formed. *FF(a man)* cannot delete against the infinitival T, hence available for further raising. Subsequently, they raise to [Spec, AgrOP] to check off the ECM (accusative) Case. The derivation converges. (67d) is bad because the insertion of *there* is unmotivated, since there is no embedded T to attract *FF(a man)* – hence no insertion.

Consider finally, the paradigms that I focused on in section 3.2.2, to illustrate the globality of the analyses reviewed there. Let’s see how the proposed T system accounts for these facts and does so in strictly local computation. The paradigms are repeated here for convenience:

- (68) (a) There seems to be a man in the room.
 (b) ★There seems a man to be in the room.
- (69) (a) ★Mary believes to be a man in the room.
 (b) Mary believes a man to be in the room.

In the course of the derivation, the following TP is formed:

- (70) [_{TP} to be [a man in the room]]

The T feature must be satisfied. Two choices arise: (a) to insert *there* from the numeration, or (b) to raise *a man*. As noted above, *there* cannot check the EPP

⁹Note that one set of features can satisfy strong T feature multiple times, through cyclic raising:

- i. a. John is believed t to t like Mary.
 b. There is believed t to be a man in the room.

This is so because Case features cannot check against tense-less T heads. In (ia), the whole category *John* successively raises to satisfy both the embedded and the matrix EPP features. In (ib), it is the complex [*there, FF(a man)*]. However, once FF checks against tensed T, it gets Case and becomes inert for further movement.

feature by itself alone. If (a) is taken, it must be accompanied with *FF(a man)* raising. The Minimality Condition (18) demands that when a number of competing operations are available (satisfying local checking (16)), the most economical one should be taken. The cost associated with (a) actually consists of two parts: $\text{cost}(\text{Merge}) + \text{cost}(\text{FM}) = 0 + \text{cost}(\text{FM}) = \text{cost}(\text{FM}) < \text{cost}(\text{overt raising}) = \text{cost}(\text{option b})$. Therefore, (a) is taken forming (71) and (68b) is banned:

(71) [_{TP} [there, *FF(a man)*] to be [_{SC} a man in the room]]

What follows next is straightforward: *seems* is merged with (71), and the matrix T attracts the complex [*there, FF(a man)*]. Although the NP subject remains in the non-Case position of the SC of (71), its features have raised to the finite T in the matrix clause and hence get Case.

Let's now turn to (69). Again, the following TP is formed:

(72) [_{TP} to be [a man in the room]]

In this case, *Mary* cannot be inserted into [Spec,TP] to satisfy the EPP. [Spec,TP] is not a theta position so that NPs without a θ -role can not occur in such positions because movement into θ -positions is in general unavailable, following Hale and Keyser (1993), Chomsky (1995). This constitutes an example of the computational trick to reduce complexity.¹⁰ Hence *Mary* is not inserted and (69a) cannot be derived.¹¹ The

¹⁰This analysis carries over to *it*-expletives, because we have assumed that *it* is an argument (Chapter 3: footnote 5). Consider the paradigm in Chomsky (1995: p347):

- i. it seems that [someone was told t [that IP]].

At an earlier stage of the derivation, we have:

- ii. [γ was told someone [that IP]]

Since the numeration contains *it*, Chomsky (in Chapter 4) assumes that the computation has two choices to satisfy the embedded EPP: either to insert *it* or to raise *someone*. Insertion of *it* is cheaper but eventually leads to divergence, for multiple reasons (e.g. the Case of *someone* is not checked). Hence, there was concluded that economy is overridden by convergence and *someone* is raised. This (global) analysis is not necessary here, as we have a handy trick: since *it* is an argument, it cannot be inserted into [Spec,TP], a non- θ position. The only option for the computation is to raise *someone*. See also section 6.1.

¹¹Note that, however, analyses in the MPLT and Collins (1997) cannot be saved by this restriction on theta positions. Other improprieties such as N-to-D raising and optionality (hence complexity)

only option available to the derivation is the overt movement of *a man*, leading to (69b)

In summary, the proposed MFA framework and the T system have resolved a number of conceptual and empirical problems in previous other analyses: optionality, complexity, the divorce between Case and agreement, optionality, complexity, etc. Furthermore, computation is strictly local and minimal, in contrast to the globality and intractability of the analyses reviewed in Chapter 3. In Chapter 5, I will show that these proposals, largely forced by the conceptual motivation to reduce complexity, has other desirable empirical consequences as well.

still remain. This is because of their analysis of the EPP: specifically, they assume that the D feature of T can be satisfied by the pure expletive *there*. I have argued that expletive insertion to [Spec,TP] must be accompanied by feature movement of the associate.

Chapter 5

Null and Displaced Subjects

This chapter explores some consequences of the proposal presented so far. According to the new conception of T (61), the formal features of subject always raise to [Spec,TP], with or without pied-piping its PF features. I will consider a number of constructions that involve the “structural subject” (in [Spec,IP(TP)]), a postulation that is largely forced by the traditional notion of EPP (Chomsky 1981). Building on the proposals in Chapter 4, I will show that a unified analysis falls out naturally in the Minimalist framework.¹ The discussion is selective and suggestive, partly because the proposed theory, if on the right track, calls for a reformulation of many problems what have been standardly assumed. Much work lies ahead.

5.1 Null Subjects

In some Romance languages and others, overt subject could be missing in a tense clause:

(73) Baila bien. (Spanish)
 dances(3sg) well.

¹I should note that such attempts to unify inverted (displaced) and null subjects have been pursued before, most systematically by Shlonsky (1987) under the Government and Binding framework. For instance, a central hypothesis for Shlonsky is the LF expletive-replacement analysis of Chomsky (1986), which has been replaced by feature movement in current work. There are other crucial differences that I will not enumerate here. As a result, different conclusions were reached.

(He) dances well.

Another important fact is subject inversion, in both unaccusatives (74) and in transitives (75):

(74) (a) E arrivato Gianni.
is arrived Gianni.
“Gianni has arrived”.

(b) Ha telefonato sua moglie.
has telephoned your wife.
“Your wife has telephoned”

(75) ha mangiato un dolce il ragazzo.
ate a cake the boy (Belletti 1988)

A traditional account for this language variation involves a parameter *pro-drop*, which specifies whether a language allows null subject. For Italian/Spanish, the value is True; for English, it's False. The postulation of *pro* in [Spec,IP] is largely a result of the Extended Projection Principle (Chomsky 1981), which requires [Spec,IP] to be structurally filled. Within the proposed theory, the traditional notion of EPP has been eliminated, in favor of the T system and its PF parametric properties. Hence the necessity of *pro* per se is obviated. This calls for a reanalysis of the null subject languages.

To start, I follow Rizzi (1986) and Jaeggli and Safir (1989) to distinguish two problems with respect to the null-subject phenomena: Licensing and Identification. In the words of Rizzi (1986: p518), licensing refers to “the conditions that formally license the null element (the conditions that allow it to occur in a given environment)”, and identification (interpretation) refers to “the way in which the content of the null element (minimally, the ϕ -features) is determined, or ‘recovered’, from the phonetically realized environment”. Identification can also be seen as conditions by which languages allow null subjects. I adopt Jaeggli and Safir’s Uniformity Condition (1989: p29-30), applicable to an array of typologically diverse languages:

- (76) (a) Null subjects are permitted in all and only languages with morphologically uniform inflection paradigms.
- (b) All inflectional paradigm **P** in a language **L** is morphologically uniform iff **P** has either only underived inflectional forms or only derived inflectional forms.

I will focus on the licensing condition. First, I assume a version of the Theta Criterion: each argument is assigned one and only theta role, and each theta role is assigned to one and only one argument – irreducible, as it seems. Thus, even for sentences like (73) “(he) dances well”, an argument must be present somewhere as thematic subject. Presumably, it is in the numeration, as a bundle of formal and semantic features. However, they still need Case to be legitimate entities, following Chomsky and Lasnik’s Case condition (1993). Consider the nature of T, proposed in (61), repeated here for convenience:

(77) The Nature of T:

- (a) T universally attracts the closest nominal features.
- (b) T is parametrically strong/weak.

A language typology is now in place. Universally, the formal features of the subject move into [Spec,TP] to check nominative Case. Languages vary with respect to (77b):

Language	Strength of T
English	Strong
Italian	Optional
Irish	Weak

The optionality for Italian probably derives from the morphological uniformity of such languages, that the content of thematic subjects can be recovered without being pronounced. Note that this optionality is different from the kinds of optional operations we rejected earlier. Consider an example:

- (79) (a) Ha parlato.
 has (3sg) spoken.

- (b) Lui ha parlato.
 He has (3sg) spoken.

Both (79a) and (79b) are convergent derivations, but they do not compete with each other. As mentioned earlier in footnote 13 of Chapter 2, they come from two **different** numerations. For (79a), the numeration N_a contains a feature bundle that specifies the formal/semantic properties of the subject (third person, male, etc.). For (79b), the numeration N_b contains, in addition, the overt PF features *lui* as well. In Italian, the optionality of the strength of T is the very reason that the computation allows N_a to converge: the T feature in N_a is weak, as allowed in such a language. For English, T is **always** strong, thus the counterpart of N_a will not converge. In this sense, the “optionality” is really an optionality for **the numeration**, not the computation.

A number of predictions immediately bear out. First, Italian can optionally drop subjects because [Spec,TP] needs not to be overtly filled. The subject of an unaccusative verb (74) can stay in-situ (direct object position, following Burzio (1986)), since only feature movement to [Spec,TP] suffices for Case checking, no PF category is required.² Hence, the representation of (73) is

- (80) [TP [FF(he)]_t [VP t baila bien]].

For (74), it is:

- (81) [TP [FF(Gianni)]_t [VP arrivato [PF(Gianni)], t_{FF}]

The so-called “subject inversion” (75) is not necessarily (rightward) inversion of the subject (Chomsky 1981, Rizzi 1982). I believe it is possible to state the phenomena as a kind kind of object shift.³ The subject can afford to stay in the post-verbal position because its features can raise to [Spec,TP] without the PF feature, an option allowed in the Italian T system.

²Contra, for example, Belletti (1988), who proposes that post-verbal subject adjoins to VP to receive nominative Case.

³See Zubizarreta (1997) for arguments from the perspective of focus and prosody. See section 5.2.2 for further discussion.

This analysis also directly explains Rizzi’s conclusion that *pro* must be governed by X_0 , and has the feature specifications of X_o (1986). Here, $X = T$. In the current analysis, *pro* is a bundle of features that raise from the subject, hence it must be licensed (checked) by T. Overt agreement between the verb (T) and the post-verbal subject is accounted for immediately, exactly as in expletive constructions (section 4.1).

This new look at *pro-drop* also explains why null subject languages permit null expletives (e.g. Jaeggli and Safir 1989; p19). Note that in the current analysis, there is no such thing as “null expletive”, just as there is no such thing as *pro*: when subject “drops”, [Spec,TP] is filled with **invisible** formal/semantic but not phonetic features. Null subject language permits null expletives because it does not have a need for it. (Overt) expletives are only needed for languages such as English, where [Spec,TP] is obligatorily filled.⁴

I will end this discussion with a few words on the acquisition of subjects. In recent years, there has been a converging view that the acquisition of subject hinges upon the development of INFL (Tense), e.g. Rizzi (1994), Wexler (1994, 1995), and Hyams (1996). This departs from the setting of the pro-drop parameter in earlier studies (Hyams 1986). Note that the T system developed here provides the grammatical basis for this line of proposals. The notion of “structural subject” reduces to a parametric PF condition that if structural subject shows up phonetically, it must appear in

⁴Incidentally, the so-called “quirky subject” phenomenon in such languages receives a natural explanation as well. Consider the following Icelandic examples from Belletti (1988) (A stands for inherent accusative; D stands for dative):

- i. Snjóa(A) leysir á fjallinu.
snow melts on the mountain.

- ii. Hann telur stúlkunni(D) hafa svelgtz á súpunni.
he believes the girl to have misswallowed on the soup.

As suggested by Belletti, quirky subject manifests the possibility of combining inherent and structural Cases. In (i), *snow* bears an inherent accusative Case selected by the verb, but also combines with nominative Case from T (INFL); in (ii), *the girl* combines inherent dative Case with accusative Case from the ECM-type verb *telur* (*believes*). The mechanisms at work is the same as we have assumed all along: the strong T feature in Germanic languages that demands [Spec,TP] be overtly filled. In those examples, it is not only possible for inherent Case bearers to raise to structural Case positions, it is **necessary**.

the nominative Case position [Spec,TP]. The proposed T system analysis supports, say, Hyams's recent proposal of underspecification of functional categories (1996). Specifically, I assume that nominative Case checking hinges on the development of T: nominative Case checking is fully functional only when T is fully developed. Because the PF condition is a peripheral property of T (77), I assume that mastery of (77b) depends on that of (77a), hence the full specification of T. The optional infinitive stage considered by Wexler (1994) is presumably a stage when the child has only fragmentary knowledge of the yet underdeveloped T system.

This directly explains the correlation between proper use of verbal inflection and subjects, originally noted by Guilfoyle (1984). The strongest prediction would be that children never use null subject with finite verbs. This is not entirely true. There are two kinds of null subjects in child language. One kind is associated with nonfinite verbs; the other represents some kind of Topic-Drop (Hyams and Wexler 1993) or Diary-Drop (Haegeman 1990) that might occur with finite verbs. An expectation of this is that null subjects co-occur with nonfinite verbs more frequently than finite verb, because the latter should not (at least more unlikely) allow null subjects. This is borne out with statistical analysis; see Poeppel and Wexler (1993) for German, Haegeman (1996) for Dutch, and Pierce (1992) for French.

The developmental model of T also predicts that proper use of expletives necessarily co-occur with lexical subjects (Hyams 1986). This is because under the feature movement framework and the T system, expletives and structural subject are the same grammatical phenomenon: in both cases, the formal features of the thematic subject raise to T. Only when T is fully developed, children begin to use the Case system properly; only when the Case system is fully in place, the PF condition for an overt category in the nominative Case position is properly learned. In this sense, expletives could still serve triggering evidence for learning the PF "subject" condition, as originally proposed in Hyams (1986).

Also relevant is the acquisition of the V2 property, which strongly correlates with that of subjects (Clahsen 1986). We assume that in V2 languages, V raises to C and [Spec,CP] is the nominative position (more on this section 5.2.2). Once the child has

learned this, and also has fully developed the T system, Case checking is in place. Since Germanic languages require an overt category in Case position ([Spec,CP] here), the child stops dropping the subject. By this time, children should also know that literally anything can fill [Spec,CP].

All in all, from grammatical analysis to acquisition, it is possible to conclude that the so-called null subject is not an independent property of the grammar, but follows from the properties of the T (Case) system.

5.2 Displaced Subjects

5.2.1 Locative Inversion

Recall that in section 3.1, we noted that locative inversion in (82) closely patterns with expletive constructions, calling for a unified analysis:

(82) Down the hill came a man.

A number of questions are at stake. Most importantly, we must determine where the locative PP is situated in (82). There are some reasons to believe that the relevant position is [Spec,TP], the structural subject position. Bresnan (1994) argues that locative PP is in fact the subject. She draws evidence for the that-trace effect, characteristic of subjects:

(83) (a) in these villages we all believe [e can be found the best examples of this cuisine].

(b) *in these villages we all believe [that e can be found the best examples of this cuisine] (modified from Bresnan 1977).

Locative PP also undergoes subject raising:

(84) (a) Near the fountain seem to have been found two purple bananas and a peach.

Near the fountain are likely to be found two purple bananas and a peach.

Near the fountain are found to be situated two purple bananas and a peach.

(Postal 1977: 148)

(b) On that hill appears to be located a cathedral.

In this village are likely to be found the best examples of this cuisine.

(Bresnan 1994: 96)

Lastly, locative inversion is allowed in embedded clauses, as observed by Hoekstra and Mulder (1990: 32):

(85) (a) We all witnessed how down the hill came rolling a huge baby cabbage.

(b) We suddenly saw how into the pond jumped thousands of frogs.

If the Specifier position of the embedded CP is already occupied by *how*, then the locative PP must be situated in [Spec,TP]. However, Jang (1997), citing personal communication with Noam Chomsky, notes that this argument is weakened if we allow CP-recursion (e.g. Authier 1992), which creates additional sites in [Spec,CP].

These arguments suggest that locative PP exhibits some subject-like behaviors. I assume without further argument that the locative PP is in [Spec,TP] at some stage of the derivation. A bit later on, I will give some arguments that the PP **must** eventually end up in [Spec,TP], forced by economy considerations.

Given the proposed MFA (49) and T (61), locative inversion can trivially be unified with expletive constructions. In locative inversion, the PP moves to [Spec,TP] to satisfy the overttness requirement of the Strong T(61b). The formal features of the subject raise, as always, but without pied-piping the PF features. Agreement between the verb and the subject follows at once, as in expletive constructions. This analysis avoids the stipulation that the D-feature of the DP (e.g. *the hill* in (82)) enters into a checking relation with T from the complement position of the PP. It is not necessary here, because we do not assume the version of the EPP that requires T to attract a D-feature.

The this proposal also explains the following contrast:

(86) (a) ★To John was spoken.

- (b) To John was given a book.

In (86a), the potential Case checker *John* has checked Case within the PP, hence nothing could possibly check off the Case feature of T – the derivation crashes. In the considerably better (86b), again, the PP satisfies (61b), but the raising of FF(*a book*) to T is available. The derivation converges.

Economy considerations make an interesting prediction: in locative inversion, [Spec,TP] **cannot** be the ultimate landing site for the PP, which means that locative inversion must be triggered by something other than the overttness requirement in (61), say, Topic or Focus. Let's see why this is so. Consider the following paradigm (87):

- (87) (a) There is a man in the room.
(b) A man is in the room.
(c) In the room is a man.

Assume that the underlying structure for (87) is the small clause (88):

- (88) [TP is [_{sc} a man in the room]]

The derivational costs of (87)(a–c) are listed in (89):

- (89) (a) $\text{cost}(\text{Merge}) + \text{cost}(\text{FM}) = 0 + \text{cost}(\text{FM}) = \text{cost}(\text{FM})$
(b) $\text{cost}(\text{OM})$
(c) $\text{cost}(\text{OM}) + \text{cost}(\text{FM})$

Clearly, $\text{cost}(87a) < \text{cost}(87b) < \text{cost}(87c)$. This implies that if *there* is available in the numeration, the subject **never** overtly raises. Furthermore, the strong T feature in English cannot trigger the movement of PP, because direct raising of the subject is **always** cheaper than and therefore blocks locative inversion. Therefore, locative inversion must be triggered by something else like Topic or Focus. It implies that locative PP only passes through [Spec,TP] at an intermediate stage of the derivation and has to move further, possibly to [Spec,CP] (cf. Bresnan 1994).⁵

This prediction seems borne out in the following:

⁵A caveat here. Note that PP fronting is always possible as topicalization:

- (90) (a) [How many stars] $_{\alpha}$ do you believe [CP t_{α} are in the sky]?
- (b) [How many stars] $_{\alpha}$ do you believe [CP there are t_{α} (in the sky)]?
- (c) \star [How many stars] $_{\alpha}$ do you believe [CP in the sky are t_{α} (at night)]?

(90b) is fairly good, if slightly worse than (90a). (90c) is, however, completely bad. The contrast between (90b) and (90c) can be accounted for as follows. In the embedded clause of (b), *there* is in the subject position [Spec,TP]. [Spec,CP] is vacant, which provides α with an intermediate landing site for movement. In the embedded clause of (90c), if the PP is situated in [Spec,TP], we would expect roughly the same judgment: α moves through the embedded [Spec,CP]. But this is clearly wrong. If the locative PP moves into [Spec,CP] from [Spec,TP], triggered by, say, [-Topic] of the embedded clause, the ungrammaticality of (90c) falls out as a Movement violation. The underlying structure of (c) is presented in (91)

- (91) \star [How many stars] $_{\alpha}$ do you believe [CP [in the sky] $_{\beta}$ [C [TP t_{β} are t_{α}]]]?

Since the embedded [Spec,CP] is occupied by β , the movement of α directly into the matrix CP is illicit.

The current analysis also explains Bresnan's conclusion (1994) that the locative PP is the subject. Here the locative PP is only a superficial subject, since it does, at some stage of the derivation, situate in [Spec,TP] – the subject position. It moves to satisfy the overt requirement of strong T. The real subject raises its features to [Spec,TP] to check off Case. We can further assume that the locative PP and the formal features of the thematic subject form a complex, as in expletive constructions

-
- (i) a. In the room John kissed Mary t.
 b. In the room there is a man t

In (i), the locative PP directly into the Topic position, say, [Spec,CP] from base position. For the derivation of (ib), my claim is that *there* is inserted to [Spec,TP] together with raising of *FF(a man)*, **before** the topicalization of *in the room*. Now an interesting but separate question arises. How do we rule out (ii):

- (ii) ?? In the room₂ a man₁ is [t_1 t_2]

I will leave this for future research. Thanks to Julie Legate and Carolyn Smallwood for discussions on this.

(though not for FI reasons; see Chapter 4 footnote 9). The complex, carrying the formal features of the “associate”, functions as the subject: raising it across *that* induces the that-trace effect (83), and raising it to matrix TP checks Case, as in raising (84).

5.2.2 Other displaced subjects

Let’s turn our attention to a few more cases where the structural subject position is filled by non-thematic arguments.

To set stage for discussion, first consider the phenomenon of Object Shift that has received significant attention in the Minimalist literature (Bobaljik and Jonas 1993, 1996, Collins and Thráinsson 1996, Jang 1997 to name a few). The paradigm example is given in (92):

- (92) Jon las bækurnar ekki.
 John read the books not.
 John didn’t read the books

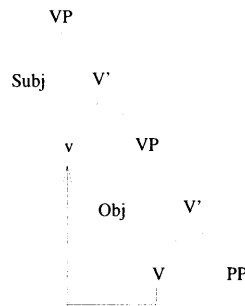
The object is shifted to the left, across the negation. Holmberg’s generalization (1986) is relevant here: object shift depends on verb movement, that is, the object NP can move into [Spec,AgrOP] external to the VP, upon the verb movement to T. Object Shift is possible because V-to-T raising renders the subject and the object equidistant. According to the typology in (78), *John* must raise overtly into [Spec,TP], since Germanic languages have a strong T feature, which must demands [Spec,TP] be overtly filled. The object, already checked Case at [Spec,AgrOP] is inert and can move no further.

Consider next the so-called Passivized Expletive Constructions:

- (93) (a) There has been a book put on the table.
 ★There has been put a book on the table.
 (b) There was a man shot in the park.
 ★There was shot a man in the park.

Lasnik (1995) accounts for these data based on Belletti's partitive Case (1988). However, the partitive Case analysis is undesirable for a number of reasons (see section 3.2.2). A solution is available, if we adopt a neo-Larsonian VP shell (Chomsky and Lasnik 1993, Chomsky 1995). Specifically, we assume the following VP structure:

(94)



The upper verb is a light verb. Adopting Larson's insights (1988) into the current framework, in transitive clauses, the lower verb raises to the light verb, assigning θ -role to the subject.⁶ When passivized as a participle, the subject role and tense are absorbed, hence the main (lower) verb stays in-situ. This directly accounts for the minimal pairs in (93). In both cases, the direct object, which is in fact the "subject" of the matrix verb *be*, raises its features to [Spec,TP]. The expletive *there* is inserted into [Spec,TP] to satisfy the strong T feature.

Consider lastly a problematic case, the so-called Transitive Expletive Constructions (TEC), mentioned earlier in Chapter 3 footnote 5. For simplicity, I use the English glosses (Chomsky 1995: p341) of the original examples in Icelandic:

(95) [AgrP There painted [TP a student t_T [AgrP [the house VP]]]]

The pre-VP positions are motivated by placement of adverbials and negations in the overt form. The expletive *there* is in [Spec,AgrSP], the subject *a student* moves into [Spec,TP], and the object *the house* moves into [Spec,AgrOP]. The VP in (95) contains only traces. Case and agreement for the object are checked overtly in [Spec,AgrOP]; for the subject, covertly [Spec,AgrSP].

⁶And moves further to T to license nominative Case.

Bobaljik and Jonas (1996) have argued that Icelandic has two positions for subjects: one ([Spec,TP]) for the thematic subject *a student* in (95), and the other ([Spec,AgrSP]) for the expletive. They assume that for languages that license [Spec,TP], TEC is possible. Icelandic is such a language, but English is not. However, they were not concerned with economy issues and did not discuss the temporal sequence of the derivation. Specifically, after the object shift, V raises to T; now the question is, what triggers the movement of the subject? The following pattern would be formed, which is bad⁷:

(96) \star [_{TP} There [_T painted [_{AgrOP} a house [_{VP} a student t_v t_{obj} .]]]]

Note that for Icelandic, expletive construction is possible: the thematic subject in unaccusatives could remain in-situ with [Spec,TP] filled by an expletive (Bobaljik and Jonas 1993). For expletive constructions, they assume that the nominative Case checking takes place at LF, following Chomsky (1995). But why doesn't the computation insert *there* into [Spec,TP] in TEC (95), as in expletive constructions, which yields (96)? I see nothing in their analysis to ban this operation.

The problem is complicated by the V2 property of Germanic languages. I follow the long-standing proposal that the finite verb raises to C in matrix clause (den Besten 1977, Holmberg 1986, Vikner 1994, to name a few). Under the proposed T system, nominative Case would be checked in [Spec,CP] upon T-C raising (cf. Koster and May 1982, Koster 1986 among others). This immediately explains that in V2 languages, almost anything can, and something must fill into the nominative Case position [Spec,CP]. If [Spec,TP] requires an overt category as attested in TEC, after V raising to T, the expletive would be inserted, with the features of the subject raising:

⁷Structures like this are actually attested, as noted by Collins (1997: p18-19), citing Bobaljik and Jonas (1993):

- i. a. There put the butter someone in the pocket
- b. There told Sveinn somebody a story
- c. There told Sveinn a student a story.

The English counterpart is marginal, though not completely bad (Chomsky, p.c.):

- ii. There hit the target a bullet.

See Chomsky (1995: p343-344) for a possible explanation.

(97) * [TP [there, FF(a student)] [a house [VP a student t_v t_{ahouse}]]]

I do not have an answer for this and will leave it for future research. Youngjun Jang (1997) made a similar point, and proposed a plausible solution. The reader is referred to his work for details.

Admittedly, this chapter is a bit speculative and I have omitted certain details. I do believe these ideas are on the right track, as an attempt to collapse cross-linguistic and construction-specific phenomena under one unified analysis – specifically, **one** parametric PF condition peripheral to the T system – natural to the conceptual motivations of the Minimalist Program.

Chapter 6

Other Minimal Properties

I will briefly consider two other applications of the proposed Minimalist theory. The first problem is the Theta Criterion, an LF interface condition, direct reference to which introduces immense complexity, as noted earlier in Chapter 2. The second is the deduction of cyclicity. These problems are of interest because they seem to pose complexity problems to the Optimality Hypothesis, and also because some recent analyses involve global computational principles. I will propose two possible local solutions, with in mind the reduction of complexity and the localization of interface conditions.

6.1 Interpretability and Economy

An important question in the Minimalist framework has to do with thematic relations (expressed at D-structure in traditional theories) and how they relate to computation.

Chomsky (1995: section 4.6) suggests the following. First of all, he adopts Hale and Keyser's (1993) configurational approach to thematic relations: θ -relatedness is a property of "base positions" under certain structural relations, complementary to feature checking, which is a property of movement.¹ Furthermore, Chomsky assumes that failure to satisfy the Theta Criterion violates FI, and thus the derivation crashes at LF.

¹See section 4.3 for the application of this distinction.

Principles such as the Shortest Derivation Requirement (SDR) figure into play. Take a concrete example. Consider the derivation of *John left*, using the notations in Chomsky (1995):

- (98) (a) Merge(*John*, *left*) = VP = {*left*, {*John*, *left*}}
- (b) Merge(T, VP) = TP = {T, {T, VP}}
- (c) Raise *John* to [Spec,T] to check case and EPP: [*John* [T t *left*]]

But what bans the following derivation which is actually shorter?

- (99) (a) Merge(T, *left*)
- (b) Merge *John* into [Spec,T] to check case and EPP: [*John* [T *left*]]

Chomsky (1994) suggests that the SDR selects among convergent derivations: shorter derivation blocks longer ones unless it does not converge. Although the derivation in (99) is shorter than that in (98), it violates the theta criterion at LF (*John* does not receive a theta role). Thus, the “ban” is lifted – (98) is pursued which yields a convergent derivation. In this sense, convergence overrides economy, and economy only chooses from convergent derivations.

This analysis violates local economy in two ways. First, it involves transderivational comparisons through SDR. Secondly, it assumes that the Theta Criterion is *only* invoked at LF. Therefore, the computation would not be able to notice that a wrong choice has been made until the entire derivation has completed and reached LF. This leads to computational complexity. Concretely, assume that for *John left*, two derivations are allowed to proceed in parallel. Consider the example in (11), repeated here as (100):

- (100) [*Carol left*] before [*Cindy left*] after [*Stephanie left*] because [*I left*].

As mentioned earlier in section 2.3, (100) requires $2^4 = 16$ derivations to proceed in parallel. Alternatively, if the derivation operates sequentially, certain space in the “memory” must be allocated to store the information to facilitate possible back-

track. This again causes combinatorial explosion. It is easy to see that the growth is intractable.²

Recall that the very essence of the Optimality Hypothesis is to localize interface conditions to reduce computational complexity. Therefore, the interpretability condition (FI) should be localized as well. In fact, we have already done so, by assuming that θ relations are base relations and can only be established by Merge. When the computation builds thematic relations in base positions, it simply looks up in the numeration for θ -role bearing categories (e.g. nouns). Since Move only affects feature checking but not theta relations, no future steps of the derivation will affect thematic structures. In particular, a θ -recipient cannot move elsewhere to receive another theta role. If no θ -role bearer is available in the numeration, the derivation crashes immediately, because we know that since theta relations cannot be established by Move. If there is nothing to merge with, the derivation **will** crash as a theta violation. No direct reference of the theta criterion is required, only local economy conditions are obeyed. This echoes Chomsky’s remark (1993: p33) that the computation is a mechanical system not in “the search for interpretability”. Interpretability at LF (convergence) should follow from local economy, instead of overriding it.

6.2 Cyclicity and Generalized Transformations

Let’s then look at the notion of strict cyclicity and its place in the computation. Chomsky’s Extension Requirement (ER) (1993) states, informally, that a syntactic operation OP must expand the structure containing OP’s target. It is used to rule

²Recall that in the derivation of (38a), *Mary_i believes t to be a man in the room*, the MPLT analysis assumes that *Mary* can be merged into the embedded [Spec,TP], for it’s cheaper than overt raising of *a man*. This leads to *Mary believes to be a man in the room*, a theta criterion violation at LF. The problem is that, there exists a convergent derivation after all, but overlooked by the computational system. Then the computation has to backtrack, an undesirable result.

Chomsky (personal communication) suggests a “local trick” to theta-related problems in general: one simply cannot merge a theta-role recipient into a non-theta position. This effectively provides a local solution (contra MPLT) for the paradigm in (38), where *Mary* cannot be merged to satisfy the embedded EPP feature and (38b) is thus banned. I have adopted this suggestion in the proposed theory (section 4.3). The objections raised in section 3.2.2 then do not arise. But see Chapter 4 footnote 11.

out the following derivation:

(101) *Who₁ do you wonder a picture of t_1 was taken?

(101) is a subject-island violation (Huang 1982). However, if the derivation takes place in counter-cyclic manner, it will be incorrectly ruled in:

(102) (a) you think [was taken [_X a picture of who]]

(b) who₁ you think [was taken [_X a picture of t_1]]

(c) who₁ you think [_{CP} [_{TP} [_X a picture of t_1] was taken t_X]]

In (102b), the attraction of *who* to the matrix [_{Spec,CP}] is cyclic - it expands the entire category and thus legitimate, but (102c) is counter-cyclic. If cyclicity (the ER) is not imposed, (102) constitutes a valid derivation – the incorrect prediction. In other words, passivization in the embedded CP must precede the wh-movement to circumvent the unwanted derivations. The question is, should the ER be stipulated or derived from more general principles?

Kitahara’s deduction of cyclicity (1995) relies on the Shortest Derivation Requirement that compares the numbers of steps to choose among derivations. As noted earlier, the SDR is a highly global principle that requires explicit and complete computation of derivations, and must be eliminated in a local theory.

Let’s see how the current theory suffices to derive the cyclicity condition, still keeping to minimal complexity. The central argument builds on a suggestion by Chomsky (1995), dubbed ASAP (19). In derivations such as (102), the Merge operation between the complementizer C and its TP complement cannot occur unless the TP has “maximally” checked off the features contained in its terms. This simply follows from the principle ASAP (19): when the TP is formed, it must check off its features immediately.³

The derivation of (101) is as follows:

(103) (a) [_{TP} was taken [_X a picture of who]]

³Also note that if ASAP (19) is true, under the mechanism of feature movement proposed here, **all** structural Cases should be checked in overt syntax, never at LF. See section 4.1.

(b) [TP [X a picture of who] was taken t_X]

At (103a), a number of features are ready to be checked off, including the strong T feature that causes nominative Case checking. Following ASAP, **something** must move to satisfy these features immediately. Thus, X in (103a) is attracted to the subject position of TP, checking off the EPP and nominative Case. Only after (103b) has been formed can the complementizer C merge with the TP to construct the embedded CP.⁴ When the entire sentence is formed, the [-Wh] feature of the matrix C attracts the [+wh] feature of *who* from the subject position of the TP in (103b), pied-piping the PF features of X required in English:

(104) *Who₁ you think [X [a picture of t_1] was taken t_X]

The movement in (104) is a standard subject-island violation, or whatever general movement condition it turns out to be.

Note that crucially, ϕ -features, and all other features irrelevant to [-wh] feature checking, are **not** attracted because they are not needed for Wh checking. The Case feature of X is already cancelled. Perhaps the following example illustrates this property better:

(105) How many stars do you believe there are t in the sky?

At the point of the derivation when the embedded TP is formed:

(106) [β there, ϕ (stars)] are [α how many stars] in the sky

It is the complex β that situates in [Spec,TP] and then moves on to check nominative Case. The category α in (106) is just a PF feature bundle plus whatever features (including [wh]) not attracted to T. When the matrix CP is formed, the [+wh] feature of α is attracted to [Spec,CP], pied-piping at least the PF features of α , because in English, [-wh] feature of C is strong:

(107) [α how many stars] do you believe [β there, ϕ (stars)] are t_α in the sky?

⁴Suppose that Wh-movement is successively cyclic, the standard assumption, overtly manifested in some languages: Spanish (Torrego 1984), Irish (McCloskey 1989), Ewe (Collins 1993), and many others.

The ASAP analysis carries over to the non-cyclicity in certain wh-island violations:

- (108) (a) *What₁ did you wonder [_{CP} how₂ Mary fixed t₁ t₂]?
(b) *How₂ did you wonder [_{CP} what₁ Mary fixed t₁ t₂]?

Before merging *wonder* with the embedded CP above, a wh word must move to satisfy the [-wh] feature of the embedded C – it cannot delay. Given the MLC, which [+wh] category should move?⁵ This depends on how we analyze “distance”, the relative position of adverbial adjuncts, and other important question. But this issue is not directly relevant to the argument here: **something** must move. At a later stage of the derivation, the longer-distance attraction to matrix [Spec,CP] in (108) is barred if it crosses a wh-island – more fundamentally an MLC violation, because the [+wh] features (categories) in the embedded [Spec,CP] are closer.

If the strict cyclicity, as derived, follows from general economy principles, it needs not to be stipulated as a condition on syntactic operations.⁶ (XP) adjunction seems to be an altogether different operation (Yang 1995, Chomsky 1996), which crucially involves no feature checking. Analyses in other grammatical frameworks also suggest that substitution and adjunction appears to be different (Frank and Kroch 1995).

The two problems considered here suggest that local computation is able to capture certain syntactic constraints without resorting to global complexity. Needless to say, we should await further evidence in favor of the local approach.

⁵Howard Lasnik (p.c) pointed out that (108a) is considerably worse than (108b). Hence the choice of the moved element is an important problem.

⁶See Richards (1997) for similar discussions and some cross-linguistic evidence.

Chapter 7

Conclusions and Speculations

I started out by presenting two conceptual motivations for a Minimalist theory: the Bare Output Conditions that determine legitimacy of syntactic entities as well as cross-linguistic variations, and the Optimality Hypothesis that assumes the derivational system of C_{HL} exhibits computational simplicity. Hence a number of **necessary** conditions are derived, if such claims are to be substantiated. They are embodied in a set of economy conditions in section 2.4, which is actually a strict subset of those proposed in the literature e.g. Chomsky (1995), with global principles completely eliminated.

7.1 Perfect Language

Let's try to give a preliminary complexity analysis for such a theory. If the local version of the Last Resort principle (16) is true, then every step of the derivation checks off (eliminates) at least one formal feature (see also Stabler 1996). Given a particular numeration N with a finite number of formal features F , the number of steps in the derivation is then bounded by $O(F)$. Suppose that at the step i of the derivation, there are m_i formal features. If all GTs are binary,¹ then at stage i , the number of derivational choices cannot exceed:

¹Merge is by definition binary. Move can be viewed as a category (feature) α raising to target β (Kitahara 1995), thus is binary as well. It is not clear to me that the operation Select has much conceptual or empirical content.

$$(109) \binom{m_i}{2} = O(m_i^2)$$

This is clearly tractable. Note that m_i is decreasing as the computation goes along, because each step eliminates formal feature(s). Therefore, $\forall i, m_i \leq F$. Since no reference is made to other stages throughout the computation, competing derivations can be safely discarded for good, once the locally optimal choice is made. The computation performs no more than $O(F \times F^2) = O(F^3)$ altogether – polynomial. Also, at any stage of the derivation, the computation only needs to keep **one** derivational path in the “memory” – characteristic of the Best-First Search (5). Both time and space requirements are polynomial to the size of the numeration (the number of formal features) and are therefore tractable. In this sense, the proposed theory achieves a kind of minimal computation.

The specific proposals in MFA (49) and the T system (61) are largely forced by conceptual motivations to reduce computational complexity. The major troublemaker is Procrastinate, which induces optionality (for Case checking) and therefore global computation. To eliminate optionality, we are almost immediately forced to the conclusion that nominative Case checking must take place in overt syntax, i.e. before Spell-Out. However, in expletive constructions, the thematic subject does remain in-situ. This leads to the idea of Minimal Feature Attraction that within one lexical item, feature relevant for checking must be available for attraction **independent** of the irrelevant features (e.g. PF). Hence, features must be “distributed” throughout the computation. The T system is conceived in a similar manner. It is observed that upon entering into the computation, expletive must immediately form a pairing relation with an associate, at the position of [Spec,TP]. Expletive’s inability to check Case suggests that its insertion into Case position must be forced by independent reasons. This eventually leads to the proposal in (61), where expletive insertion is viewed as a peripheral PF property of T, parametric across languages. Some desirable consequences follow, as presented in chapter 4 and 5, including a unification of null and displaced subjects. Conceptual motivations and empirical facts converge quite smoothly. Needless to say, a lot more has to be done to substantiate the theory

proposed here. The results presented in this thesis suggest that such an expectation might not be unreasonable.

If this Minimal Computation Theory is true, we will have some interesting evidence for the minimalist intuition and the Optimality Hypothesis. That is, the C_{HL} is indeed something like a perfect solution to its output interface conditions. It demonstrates a remarkable degree of optimality and elegance: local optimality ensures global optimality, with radical reduction in complexity.

It is frequently alleged that language is a cultural or social convention, or an artifact that arises under general intelligence that cuts across domains of knowledge. Overwhelming linguistic evidence has been accumulated against such claims. For a complexity point of view, even a cursory examination reveals that properties of language are fundamentally different from some artifactual systems that resemble the generative character of human language. For instance, consider Chess, which has induced much hype as an indication of human / machine intelligence ever since the 50s to the present day. Chess can be viewed as a search problem in strategy space. Call the computational system of chess C_{CH} . The task of C_{CH} is to take an initial board configuration (similar to the numeration N) to a specific configuration. Without loss of generality, let's take that configuration to be a winning configuration.² The analogy of the Optimality Hypothesis would demand the C_{CH} do more than just winning: It must also win in the most "economical" way, say, in the fewest number of steps. Note that this problem is computable, that is, there exists a combinatorial algorithm that enumerates all possible developments from the initial configuration and chooses the shortest path leading to victory – not much different from what Deep Blue does – it is simply impractical to run it. Indeed, this problem is provably hard – **extremely hard** – exponential-time complete (Fraenkel and Lichtenstein 1981). One cannot possibly imagine anything in the C_{CH} similar to the C_{HL} . The latter has a remarkable degree of optimality: by taking locally optimal derivation, global optimality is guaranteed. In terms of chess, this would amount to beating an opponent by blindly taking the

²Or to give a answer, "this game is unwinnable (unreachable)", similar to the interface conditions, which signal the "success" or "failure" of a derivation.

locally “best” moves (defined properly in terms of reward or loss). This, if true, would not take a two-ton computer to beat Kasparov (when he was having a bad day). Compare then the C_{HL} to a formal proof system, a generative system that manipulates abstract axioms with logical rules to form theorems. The path leading to a particular theorem cannot at all be localized or captured in any deterministic yet efficient manner; otherwise, the efforts of mathematicians since Fermat to 1993 would be instantly defied. In both chess and proof systems, global optimality can only be achieved through genuine creativity or brute-force computer power, and they just have quite different complexity properties from human language.

7.2 Imperfect Use

The Optimality Hypothesis should also shed light on the nature of language use. Parsing can be viewed as taking the PF form π as input, assigning a structural description to it, and producing the LF entity λ as output. Interestingly, complexity analysis comes into play again. A parse can be viewed as the output of a derivation; parsing then amounts to finding **the** derivation that produces the input PF form π . Suppose the theory considered here is correct and the Optimality Hypothesis is true. Then, for a given derivation D , we can instantly check if it is **the** derivation for a given π . Very importantly, just the fact that D generates π doesn’t make it the “right” derivation: in Minimalist theory, D must be the most economical among all the derivations that can potentially generate π . If the local theory is true, local optimality ensures global optimality. That is, we can check D step by step, and examine if each step is indeed the locally optimal one. If so, we **know** that D is the globally optimal thus the most economical derivation – **the** derivation. Thus, verification for a candidate parse (derivation) is easy. This immediately brings us back to the verification vs. solution game alluded to in Chapter 2. If verification for a candidate derivation (parse) is easy, as the current theory suggests, solution, i.e. finding the “right” derivation (parse) is likely to be hard. This speculation is consistent with the more technical conclusions in Barton et al (1987) and Ristad (1990). In other words, if we are not angels that

compute like nondeterministic Turing machines, or $NP \neq P$, then the human parser, in fact **any** parser, will have a hard time parsing sentences in general – which we do:

(110) # This sentence [the C_{HL} [the Optimality Hypothesis, if true, entails] generates] is unparsable.

I have implemented a parser (Yang 1995) based on some aspects of the proposed theory and Epstein's (1995) derivational formulation of c-command. The parser works by making educated guesses. By making use of local economy principles, it trims down the parsing (search) space to reconstruct derivations for the input string. The behavior of the parser is quite different from the theory of the C_{HL} it implements. For one thing, the parser **does** perform non-local computation, e.g. backtrack, and this should not come as a surprise at all. Even if the optimal theory is true, it does not imply that the use of the very theory results in an optimal or efficient parser. Quite the opposite is true. A theory is almost certainly wrong if it lends to an "optimal" parser. This is because in human sentence processing, difficulties and errors do arise. Furthermore, if Ristad's NP-completeness hypothesis is true, then the parsing human language is in general hard, for any parser, even those with psychologically implausible computational resources. This leads to the speculation that language is in general unusable (unparsable), falling out of our processing capacity. What is being used, certainly not anything like (110), is the overlap between the expressive power of C_{HL} and our (rather limited) cognitive constraints.

The preliminary work presented in this thesis suggests that the C_{HL} demonstrates an interesting kind of computational simplicity and elegant, much as the Optimality Hypothesis claims. If this line of research turns out correct, we are led to some interesting and indeed surprising discoveries into the computational structure of human language.

Auf der Straße steht ein Lindenbaum,
Da hab ich zum erstenmal im Schlaf geruht!
Unter dem Lindenbaum, de hat
Seine Blüten über mich geschneit,
Da wußt ich nicht, wie das Leben tut,
War alles, ach, alles wieder gut!
Alles! Alles! Lieb und Leid!
Und Welt und Traum!

Gustav Mahler, Lieder eines fahrenden Gesellen

References

1. Abraham, Werner, Samuel D. Epstein, and Hoskul Thrainsson (eds) (1996). *Minimalist ideas: Syntactic studies in the Minimalist framework*. Amsterdam: John Benjamins.
2. Authier, J. Marc (1992). Iterated CPs and embedded topicalization. *Linguistic Inquiry* 23, 329-336.
3. Barton, Ed, Robert C. Berwick, and Eric Ristad (1987). *Computational complexity and natural language*. Cambridge, MA: MIT Press.
4. Belletti, Adriana (1988). The case of unaccusatives. *Linguistic Inquiry* 19, 1-34.
5. Bobaljik, Jonathan and Dianne Jonas (1993). Subject positions and the role of TP. Ms, MIT.
6. Bobaljik, Jonathan and Dianne Jonas (1996). Subject positions and the role of TP. *Linguistic Inquiry* 27, 195-236.
7. Bouchard, Denis (1984). *On the content of empty categories*. Dordrecht: Foris.
8. Bresnan, Joan (1977). Variables in the theory of transformations. In Peter Culicover, Tom Wasow, & Adrian Akmajian (eds.) *Formal Syntax*. New York: Academic Press.
9. Bresnan, Joan (1994). Locative inversion and the architecture of universal grammar. *Language* 70, 72-131.
10. Brody, Michael (1995). *Lexico-Logical Form*. Cambridge: MIT Press.
11. Bures, Anton (1993). There is an argument for a cycle at LF, here. In CLS 28, Vol. 2, 14-35. University of Chicago, Chicago.
12. Burzio, Luigi (1996). *Italian syntax*. Dordrecht: Reidel.
13. Chomsky, Noam (1980). *Rules and Representations*. New York: Columbia University Press.

14. Chomsky, Noam (1977). Conditions on rules of grammar. In *Essays on form and interpretation*. New York: North-Holland.
15. Chomsky, Noam (1981). *Lectures on government and binding theory*. Dordrecht: Foris.
16. Chomsky, Noam (1986a). *Barriers*. Cambridge, MA: MIT Press.
17. Chomsky, Noam (1986b). *Knowledge of language*. New York, Prager.
18. Chomsky, Noam (1988/1991). Linguistics and Cognitive Sciences: Problems and Mysteries. In Asa Kasher (ed.), *The Chomskyan Turn*. Oxford: Blackwell.
19. Chomsky, Noam (1991). Some notes on economy of derivation and representation. In Robert Freidin (ed.) (1991). *Principles and parameters in comparative grammar*. Cambridge, MA: MIT Press.
20. Chomsky, Noam (1993). A Minimalist Program for Linguistic Theory. In Hale and Keyser (eds) (1993).
21. Chomsky, Noam (1994). Bare phrase structure. Ms, MIT.
22. Chomsky, Noam (1995). *The minimalist program*. Cambridge, MA: MIT Press.
23. Chomsky, Noam (1996). Some notes on the notion of economy in generative grammar. Ms, MIT.
24. Clahsen, Harald (1986). Verbal inflection in German child language: Acquisition of agreement markings and the functions they encode. *Linguistics* 24, 79-121.
25. Clahsen, Harald (ed) (1996). *Generative perspectives on language acquisition*. John Benjamins: Amsterdam.
26. Collins, Chris (1993) Topics in Ewe Syntax. Ph.D dissertation, MIT.
27. Collins, Chris (1994). Economy of derivation and the Generalized Proper Binding condition. *Linguistic Inquiry* 25, 46-61.

28. Collins, Chris (1997). *Local Economy*. Cambridge, MA: MIT Press.
29. Coopmans, Peter (1989). Where stylistic and syntactic processes meet: locative inversion in English. *Language* 65, 728-751.
30. Emonds, Joseph (1978). The verbal complex V'-V in French. *Linguistic Inquiry* 9, 49-77.
31. Epstein, Samuel D. (1995). Un-principled syntax and the derivation of syntactic relations. Ms, Harvard University.
32. Frampton, John (1995). Expletive insertion. Ms, Northeastern University.
33. Fraenkel, A. S. and D. Lichtenstein (1981). Computing a perfect strategy for $n \times n$ chess requires time exponential in n . *J. Combin. Theory (Ser. A)* 31: 199-214.
34. Frank, Robert and Tony Kroch (1995). Generalized transformations is successive cyclic and long dependencies. Ms, UPenn.
35. Fukui, Naoki (1993). Parameters and optionality. *Linguistic Inquiry* 24, 399-420.
36. Gary, Michael R. and David S. Johnson (1979). *Computers and intractability: A guide to the theory of NP-Completeness*. W. H. Freeman: New York.
37. Guilfoyle, Eithne (1984). The acquisition of tense and emergence of lexical subjects. Ms, McGill University.
38. Groat, Erich (1995). English expletives: a minimalist approach. *Linguistic Inquiry* 26, 354-365.
39. Hale, Kenneth and Samuel Jay Keyser (1993). On argument structure and the lexical expression of syntactic relations. In Hale and Keyser (eds) (1993)
40. Hale, Kenneth and Samuel Jay Keyser (eds) (1993). *The View from building 20*. Cambridge, MA: MIT Press.

41. Halle, Morris and Alec Marantz (1993). Distributed morphology and the pieces of inflection. In Hale and Keyser (eds.) (1993).
42. Halle, Morris and Alec Marantz (1994). Some key features of distributed morphology. Ms. MIT.
43. Haegeman, Liliane (1990). Understood subjects in English diaries. *Multilingua* 9: 157-199.
44. Haegeman, Liliane (1996). Root infinitives, Clitics, and truncated structures. In Clahsen (ed) (1996).
45. Hoekstra, Teun and Rene Mulder (1990). Unergatives as Copular Verbs: Locational and Existential Predication. *The Linguistic Review* 7, 1-79.
46. Holmberg, Anders (1986). Word order and syntactic features in the Scandinavian languages and English. Doctoral dissertation, University of Stockholm.
47. Huang, James (1982). Logical relations in Chinese and the theory of grammar. PhD Dissertation, MIT.
48. Hyams, Nina (1986). *Language acquisition and the theory of parameters*. Dordrecht: Riedel.
49. Hyams, Nina (1996). The underspecification of functional categories. In Clahsen (ed) (1996).
50. Hyams, Nina and Ken Wexler (1993). On the grammatical basis of null subjects in child languages. *Linguistic Inquiry* 24, 412-459.
51. Jaeggli, Osvaldo and Ken Safir (eds) (1989). *The Null subject parameter*. Dordrecht: Kluwer.
52. Jang, Youngjun (1997). Tense and Complementizer Feature-Checking. Doctoral dissertation, Harvard University.

53. Johnson, Kyle (1991). Object positions. *Natural Language and Linguistic Theory* 9, 577-636.
54. Johnson, David and Shalom Lappin (1996). A critique of the minimalist program. Oct 8, 1996. Ms, IBM Watson Research Center.
55. Kayne, Richard (1994). *The antisymmetry of syntax*. Cambridge, MA: MIT Press.
56. Kawashima, Ruriko and Hisatsugu Kitahara (1994). Strict cyclicity, linear ordering, and derivational c-command. In WCCFL 1994, University of Southern California.
57. Kitahara, Hisatsugu (1995). Target alpha: Deducing strict cyclicity from derivational economy. *Linguistic Inquiry* 26: 47-77.
58. Kitahara, Hisatsugu (1997). *Elementary operations and optimal derivations*. Cambridge, MA: MIT Press.
59. Koizumi, Masatoshi (1995). Phrase structure in minimalist syntax. Doctoral dissertation, MIT.
60. Koster, Jan (1986). *Domains and dynasties*. Dordrecht: Foris.
61. Koster, Jan and Robert May (1982). On the constituency of infinitives. *Language* 58, 117-143.
62. Lasnik, Howard (1981). Learnability, restrictiveness, and the evaluation metric. In C. L. Baker and John J. McCarthy (eds) *The logical problem of language acquisition*. Cambridge, MA: MIT Press.
63. Lasnik, Howard (1994). Last Resort. Ms, UConn.
64. Lasnik, Howard (1995). Case and expletives revisited: On greed and other human failings. *Linguistic Inquiry* 27, 615-634.

65. Lasnik, Howard (1995a). A note on pseudogapping. In Rob Pensalfini and Hiroyuki Ura (ed). *Papers on Minimalist Syntax*. MITWPL.
66. Lasnik, Howard (1996). On certain structural aspects of anaphora. Ms. UConn.
67. Lasnik, Howard and Mamoru Saito (1984). On the nature of proper government. *Linguistic Inquiry* 15, 235-289.
68. Lasnik, Howard and Mamoru Saito (1991). On the subject of infinitives. In Proceedings of the 27th Annual Meeting of the Chicago Linguistic Society.
69. Legate, Julie Ann and Carolyn Smallwood (1997). The Case filter meets the Minimalist Program: Evidence for strong Case features. Ms, University of Toronto.
70. Levin, Beth and Malka Rapaport (1995). *Unaccusativity: At the syntax-lexical semantics interface*. Cambridge, MA: MIT Press.
71. Longobardi, Giuseppe (1994). Reference and Proper Names. *Linguistic Inquiry* 25, 609-666.
72. McCloskey, James (1991). *There, it, and agreement*. *Linguistic Inquiry* 22, 563-567.
73. Milsark, Gary (1974). Existential sentences in English. Doctoral dissertation, MIT.
74. Milsark, Gary (1977). Towards an explanation of certain peculiarities of the existential construction in English. *Linguistic Analysis* 3, 1-31.
75. McGinnis, Martha Jo (1995). Fission as feature movement. Ms. MIT.
76. Pierce, Amy (1992). *Language acquisition and syntactic theory: A comparative analysis of French and English child grammars*. Dordrecht: Kluwer.
77. Poeppel, David and Ken Wexler (1993). The full competence hypothesis of clause structure in child German. *Language* 69, 1-23.

78. Pollock, Jean-Yves (1989). Verb movement, universal grammar, the structure of IP. *Linguistic Inquiry* 20, 365-424.
79. Prince, Alan and Paul Smolensky (1993). The Optimality Theory. Ms, Rutgers University.
80. Poole, Geoff (1995). Constraints on Local Economy. In Barbosa et al (eds) *Is the best good enough* (to appear). Cambridge, MA: MIT Press.
81. Postal, Paul (1977). About a “non-argument” for raising. *Linguistic Inquiry* 8, 141-154.
82. Richards, Norvin (1997). Featural cyclicity and the ordering of multiple specifiers. Ms, MIT.
83. Ristad, Eric (1990). Computational structure of human language. Ms, MIT AI Lab.
84. Rizzi, Luigi (1982). *Issues in Italian syntax*. Dordrecht: Foris.
85. Rizzi, Luigi (1986). Null objects in Italian and the theory of pro. *Linguistic Inquiry* 17: 501-557.
86. Rizzi, Luigi (1990). *Relativized minimality*. Cambridge, MIT Press.
87. Rothstein, Susan (1983). The syntactic form of predication. Ph.D dissertation, MIT.
88. Safir, Ken (1982). Syntactic chains and the definiteness effect. Doctoral dissertation, MIT.
89. Safir, Ken (1985). *Syntactic chains*. Cambridge: Cambridge University Press.
90. Sauerland, Uli (1996). The late insertion in Germanic inflection. Ms. MIT.
91. Shlonsky, Ur (1987). Null and displaced subjects. Doctoral dissertation, MIT.

92. Stabler, Edward (1996). Acquiring and using language with movement. Ms, UCLA.
93. Torrego, Esther (1984). On Inversion in Spanish and some of its effects. *Linguistic Inquiry* 15, 103-129.
94. Vikner, Sten (1994). *Verb movement and expletive subjects in the Germanic languages*. Oxford: Oxford University Press.
95. Ura, Hiroyuki (1996). Multiple feature checking: A theory of grammatical function splitting. Doctoral dissertation, MIT.
96. Watanabe, Akira (1992). Subjacency and S-Structure movement of wh-in-situ. *Journal of East Asian Linguistics* 1, 255-291.
97. Watanabe, Akira (1993). Agr-based Case theory and its interaction with the A-bar system. Doctoral dissertation, MIT.
98. Wexler, Ken (1994). Optional infinitives, head movement and the economy of derivations. In David Lightfoot and Norbert Hornstein (eds.), *Verb movement*. Cambridge: Cambridge University Press.
99. Wexler, Ken (1995). Feature-interpretability and optionality in early child grammar. Paper presented at the Workshop on Optionality, Utrecht.
100. Williams, Edwin (1980). Predication. *Linguistic Inquiry* 11, 203-238.
101. Winston, Patrick Henry (1994). *Artificial Intelligence*. Reading, MA: Addison-Wesley.
102. Yang, Charles (1995). Derivational syntactic relations, feature checking, and local computation. Ms, MIT AI Lab.
103. Yang, Charles (1996). Type transparency and derivational minimalism. Ms, MIT AI Lab.

104. Zubizarreta, Maria Luisa (1997). Prosody, focus, and word order. Ms, University of Southern California.

3-3-5²