Generalized Transformations and Beyond

(Reflections on Minimalist Syntax)

by Hans-Martin Gärtner

Draft Version (6.12.2001). to appear. Akademie-Verlag: Berlin

Contents

Preface

The following study is a revised version of my dissertion. I would like to express my thanks, first of all, to my supervisors, Günther Grewendorf and Manfred Bierwisch, for support and patience as well as the inspiring intellectual climate they have been able to create and foster at Gräfstrasse and Jägerstrasse, respectively. Thanks also to the people on my thesis committee, Helen Leuninger, Eckhard Lobsien, and Marlies Hellinger.

Next, I would like to especially thank Marcus Kracht for his unfailing readiness to share his enormous knowledge and enthusiasm. Major thanks goes to John Frampton, whose input at an early stage put me on my track. Next, Reinhard Blutner, Anatoli Strigin, and Chris Wilder deserve my gratitude for providing such broad and fascinating perspectives on linguistics as well as being available for help with tricky questions.

Thanks also to the wider groups of linguists I had the pleasure to interact with and learn from. Damir Ćavar, Kathrin Cooper, Christine Erb, Michael Gamon, Katharina Hartmann, Thomas Köhn, Young-Mee Park, Joachim Sabel, Stephan Schmidt, Ruth Seibert, and Jochen Zeller in Frankfurt. Artemis Alexiadou, Kai-Uwe Alter, Assinja Demjjanow, Johannes Dölling, Gerhard Jäger, Ursula Kleinhenz, Paul Law, Annette Lessmöllmann, Christine Maassen, André Meinunger, Renate Musan, Chris Piñón, Michal Starke, and Ilse Zimmermann in Berlin. Joanna Błaszczak, Gisbert Fanselow, Susann Fischer, Thomas Hanneforth, Shin-Sook Kim, Jens Michaelis, Matthias Schlesewsky, Peter Staudacher, and Christian Wartena in Potsdam.

Further thanks goes to the wider community of linguists, GGS attendants, and guests at the ASG-Berlin, in particular, Josef Bayer, Daniel Büring, Hubert Haider, Paul Kiparsky, Gereon Müller, Jairo Nunes, Christer Platzack, Wolfgang Sternefeld, and Sten Vikner.

Very special thanks for linguistic company and friendship over the years to Markus Steinbach and Ralf Vogel.

I'm greatly indebted to the Max-Planck Gesellschaft zur Förderung der Wissenschaften, which supported my work for over three years with utmost generosity. Also to the Department of Linguistics at Stanford for hosting me during the last stages of finishing this book.

Part of these materials have been presented at the GGS-Meeting in Salzburg 1998, and WCCFL 18 in Tucson 1999. Thanks to the audiences for comments and suggestions.

I dedicate this book to my parents.

1 Introduction

At most mere minimum. Meremost minimum.
(Samuel Beckett, Worstward Ho)

Suppose you get up one morning and decide to write a dissertation. You sharpen a set of pencils and arrange them neatly next to a pile of radically blank sheets of paper. As you take a sip from your cup of tea, there is a knocking on the door, and before you have a chance of pretending that you want to be disturbed, somebody rushes in and asks you:

Q1: How “perfect“ is language?

“Not very!“ you answer after some deliberation. “Otherwise writing a dissertation wouldn't be such a demanding task, not to speak of reading one.“

Now, Q1 is the key question raised by Chomsky (1995a, p.9) in defining the “minimalist program.“ And that is what this dissertation is meant to contribute to. Of course, “language“ has to be taken in the technical sense of “I-language,“ or “competence,“ established in generative linguistics. The point of Q1, then, is to hypothetically assume that “I-language“ were the minimal device capable of relating sound and meaning, the task “language“ is taken to have to fulfill by “virtual conceptual necessity.“ And, the follow-up claim to this is that such a hypothetical assumption has major repercussions for syntactic frameworks like “Government-Binding Theory“ (GB) (cf. Chomsky 1981). In particular, the phrase structural and transformational component of GB have to be unified in roughly the following way.

Starting from a pool of lexical items, linguistic expressions are built by two operations, “binary-“ and “singulary transformations.“ The former, called “generalized transformations“ in earlier frameworks (cf. Chomsky 1975/1955[1]) and “Merge“ in current minimalist theory, applies to two objects and combines them into a larger whole. This is illustrated in (1).

(1) BT(α,β)
 $\alpha \quad \beta \rightarrow (\alpha, \beta)$

A singulary transformation, called “Move“ in minimalist theory, applies to a single complex structure α, locates a substructure β inside α, lifts β out of α, combines it with α into a larger whole, and leaves a recording “trace“ or “copy“ of β in the original position of β. This is sketched in (2).

(2) $ST((_{\alpha}\ \beta\ ))$
$(_{\alpha}\ \beta)\ \rightarrow\ (\ \beta,\ (_{\alpha}\ (\textsc{copy}\ \beta)\ )\ )$

The challenge for minimalism is to rebuild syntactic theory around these two procedures with minimal recourse to auxiliary technical devices. The way Chomsky (1995a) goes about meeting this challenge will be the core issue of the first, "reflective" part of this study (= section 2). The main thrust of this section is two-fold.

First, the "bottom-up" perspective on Merge and Move unveils a property they share, potentially leading to further unification. In graph-theoretic terms, this property consists in the constant addition of a new "root node." This is brought out by the transitions in (3) and (4), corresponding to Merge and Move, respectively.

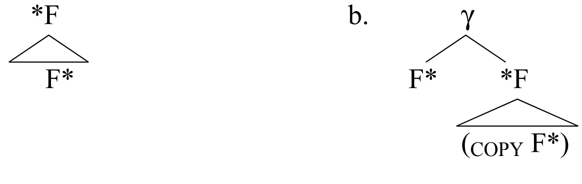
(3) a. b.


(4) a. b.


Picking up terminology from earlier frameworks, this property can be called "strictest cyclicity." In negative form it figures in one of the main hypotheses of this study, namely, H3 (cf. sections 2.5.2, 2.7.2).

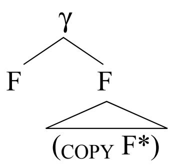
Hypothesis 3: There are no counter-cyclic syntactic operations (=H3)

Secondly, the copying part of Move is considered problematic. This has to do with the "resource sensitivity" of minimalist syntax, according to which Move can only apply if it leads to the elimination, or "checking," of certain features. Thus, take α to be a "functor," *F, and β an "argument," F*. The idea is that the two counterparts cancel against each other as soon as they are brought into local contact, i.e. [F* *F] → [F F]. As soon as these "checking resources" are eliminated, no further, "superfluous" operations can apply. If any such resources survive computation, on the other hand, the output is ill-formed. Thus, consider again a transition like (4), where *F and F* replace α and β, respectively.

(5) a. b.


Local cancelation in (5b) produces (6).

(6)

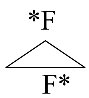


Clearly, given the survival of (COPY F*), checking resources cannot directly be fully exhausted this way. I call this the “resource paradox.“ The remedy suggested in this study is to disallow the kind of copying involved in Move. Instead, I allow constituents to be immediate constituents of more than one larger constituent, or, equivalently, nodes to be immediately dominated by more than one distinct node. This is technically expressed in hypothesis 5.

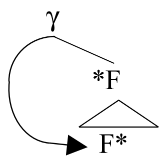
Hypothesis 5: The proper treatment of (unbounded) dependencies in minimalist syntax requires appeal to “multiconstituency“/“multidominance“ (MC/MD) (=H5)

Under this perspective, transition (7) replaces (5).

(7) a.



b.



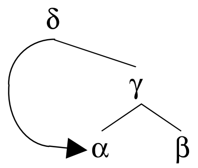
This preserves the one-to-one relationship between functors and arguments and thus voids the “resource paradox.“

The (re-)constructive part of this dissertation shows how to develop such an “MC/MD-system“ (= section 3). Its core is built around a single, hybrid, binary operation called “DoID“ (cf. 3.3.3; or alternatively “DoIC,“ cf. 3.3.1), which subsumes Merge and Move. It crucially updates an “immediate dominance“ (ID-)relation, adding one node every time. This looks roughly as follows.

(8) DoID(α,β)
$ID = \{\ \} \rightarrow ID' = \{ \langle\gamma,\alpha\rangle, \langle\gamma,\beta\rangle \}$

Multidominance is the result of applying DoID to the same argument more than once. In this respect, the system coincides with a proposal by Bobaljik (1995a). Thus, building a graph like (9) requires the steps in (10).

(9)



(10) a. DoID(α,β); ID = $\{ <\gamma,\alpha>, <\gamma,\beta> \}$
b. DoID(γ,α); ID' = $\{ <\gamma,\alpha>, <\gamma,\beta>, <\delta,\alpha>, <\delta,\gamma> \}$

There are two main reasons for why this is not a matter of a 20 page essay but a full-fledged dissertation. First, the minimalist ban on auxiliary devices deprives me of a direct appeal to graph-theory. Instead, I try to explore how much of set-theory is in principle available in standard minimalist syntax, such that it can be applied to reconstruct graph-theory without breaking the rules of the game. Unfortunately, this proves to be a “non-trivial“ enterprise, leaving me ultimately unable to tell to what extent I'm guilty of disregarding the following maxim (Chomsky 1995a, p.225).

> “In pursuing a minimalist program, we want to make sure that we are not inadvertently sneaking in improper concepts, entities, relations, and conventions.“

Secondly, the “protective belt“ of the minimalist program, constituted by appeal to “interface-“ and “economy principles,“ has to be constantly taken into account. I have tried to document at various places (cf. 2.2., 2.6.4, 2.7.1, 2.8, 3.1) why I'm skeptical of these extra devices.

Let me add three major caveats here. First, the minimalist program is still very much in flux (cf. Chomsky 2000, p.89), so no strong conclusions can be drawn from any particular criticism of any particular stage of it. This dissertation concentrates on the 1995-incarnation of the program, itself a set of various theory fragments. Meanwhile, the program has undergone further revisions, documented in Chomsky (2000, 2001). I have been unable to assess these in any close detail. A first look, however, indicates that a number of critical points raised in this study have been rendered “obsolete“ by modifications. I have tried to address some of the most crucial ones in footnotes, and, especially section 2.7.4. The least I can therefore hope is that this study contributes to an understanding of why certain modifications may (have) be(en) advisable.

Secondly, this is not an introduction to minimalist syntax. Although some ground is covered in minute, sometimes tedious and repetitive detail, other parts are only very sketchily addressed or even ignored.

Finally, although there is a certain amount of formalism provided, this reaches its limits where my own understanding ends. The least I hope to be able to convey is that formalization will ultimately be necessary in order to assess such difficult questions as Q1. The next step will have to be left to the experts.

1.1 Conceptual Background

Given the rather global nature of question 1 (Chomsky 1995a, p.9),

Q1: How “perfect“ is language?

some general remarks are in order on my part concerning the conceptual bearings of this study. It has quite regularly been observed that

> “[. . .] it is an essential prerequisite to an understanding of the development of generative grammar to have a clear picture of the aims and philosophical underpinnings of the Chomskyan programme; it is only with these in mind that the full force of the arguments for or against some technical innovation can be appreciated“ (Horrocks 1987, p.1).

Although it is tempting to broach the more philosophical subject of how knowledge of language relates to cognition in general, and overarching economy principles in particular, I will refrain from doing that here.[1] However, since the study to follow will get me involved in technical questions to a considerable extent, I would like to at least cursorily indicate why I attribute this importance to technicalities.

One central point is an emerging tendency of minimalism that can be brought out by what I will call “Frampton's conjecture.“

(11) Frampton's conjecture
Competence theory is on its way toward an algorithmic characterization of mental computations[2]

Whether mental computation taken literally is, or should in the long run be, an explicit aim of competence theories, is a difficult question. Kolb (1997a) argues that pursuing such an aim would essentially dissolve the object of study of generative grammar. Indeed, at a certain level of abstraction, generative theories have always dealt with “mental computation.“ This at least is a legitimate interpretation of the following remark by Chomsky (1991, p.5).[3]

> “The brain, like any other system of the natural world, can be studied at various levels of abstraction from mechanisms: for example, in terms of neural nets or computational systems of rules and representations.“

---

1 For discussion see for example Chomsky (1980a, 1980b, 1986a, 1991, 1995c, 1998), Grewendorf (1995), Jackendoff (1997), Kolb (1997a), Putnam (1975), and Stabler (1983), as well as contributions to Haugeland (ed.) (1981) and Macdonald&Macdonald (eds.)(1995).

2 Cf. Frampton (1997, p.40). In fact, Frampton is much more careful in his actual formulations. I'm overstating the case in order to bring out what I consider to be hidden assumptions.

3 Cf. Chomsky (2000, p.142fn.24).

Thus, Frampton's conjecture should be understood as saying that competence theory is getting closer to a characterization of (brain) "mechanisms." More precisely, it is getting closer to what David Marr called the "algorithmic level" (C.Macdonald 1995, p.293f.).[4] Now, even if this is only an implicit guideline, it may be considered disconcerting how little generative linguists seem to take seriously what's happening in neighboring fields, such as computer science. Note that computer jargon is quite pervasive throughout the Minimalist Program. "Array," "Merge," "shortest path," and "greed"(-y algorithm) are perhaps most conspicuous. These terms can be found in the index of standard introductions to computer science or theories of algorithms.[5]

Additionally, one might wonder when it was that the core object of generative research, the generative device formerly called "(universal) grammar," or "I-language," acquired the name "computational system."[6]

Innocent as all of this may sound, given the possibility of noncommittal usages, or theory-internal determination of the meaning of terms, as proposed for "simple" and "powerful" in Chomsky (1965), it remains to be seen whether or not the lack of transparency surrounding the terms that allude to computer science is welcome.

It has always been taken for granted in, and been considered one of the strengths of generative linguistics that

---

4 Someone who tackled such questions in order to find out what is at stake is Stabler (1983). If I understand that paper correctly, he claims that competence theories by nature provide descriptions at, what Marr would call, the "computational" level, i.e. where "the functions computed at the algorithmic level are described in mathematically transparent terms which abstract from the ways in which the computer carries those functions out" (C.Macdonald 1995, p.294). Frampton's conjecture would *prima facie* deny this, which makes it more provocative. The following quote from Chomsky (1971) seems to be a clear statement in favor of Stabler's interpretation. The paragraph preceding it defends the abstract nature of generative operations when defining competence against procedural interpretations that belong to performance models. "To confuse the two kinds of account would be a *category mistake*. In short, it is necessary to observe the difference in *logical* character between performance and competence" (1971, p.188; italics mine, H.M.G.). More recent remarks (e.g. Chomsky 1995a, p.380fn.3) allow for speculations of the type referred to here as Frampton's conjecture. Of course, some branches of generative linguistics have anticipated this potential drift and deal with computational issues explicitly. See Abeillé (1993) for a thorough introduction to LFG, GPSG, HPSG, and TAG grammars. For getting confused, see also the exchange between Chomsky and Schank in Chomsky (1980b).

5 Goldschlager&Lister (1988) and Cormen et al. (1990).

6 It looks as if Chomsky [with Howard Lasnik] (1995a, chapter 1) is the source. Note, however, the reference to a "system of mental computation" in Chomsky (1980c, p.1). Consider also the following assessment by Jerry Fodor, whose philosophical and psychological work has provided part of the horizon for the generative enterprise since the early days: "The available models of cognitive processes characterize them as fundamentally computational and hence presuppose a representational system in which the computations are carried out" (1975, p.99), and Chomsky's characterization of the cognitive revolution of the 1950s as crucially involving "representational-computational theories of the mind" (1991, p.4). Günther Grewendorf (p.c.) points out the difficulties the term "computational" poses when it has to be translated into German. For further clarification, alternatives, and critical assessment of what's going on in cognitive science see for example Macdonald&Macdonald (eds). (1995), Haugeland (ed.) (1981), and Winograd (1987).

> "mappings are effective - that there is an algorithm for enumerating sentences, structural descriptions, and grammars [. . .]" (Chomsky 1965, p.202fn.18).[7]

Yet, linguistic innovations since the 1960s have not generally been accompanied by interest in formal innovations achieved in neighboring fields like computer science.[8] Work on natural language processing is an exception although opinions on its progress and relevance to competence theory diverge (cf. Chomsky 1995a, Stabler 1996, and Abeillé 1993).

In any case, the renewed interest in general matters of simplicity like Q1, and the workings of economy principles within grammar in particular, threatens to fall into a theoretical vacuum unless technical detail is made explicit. To pursue a number of technical issues involved in Chomskyan minimalism in quite minute detail, including possible alternatives, is going to be one of the objectives of my study. Still, a lot of work will have to be left to the specialists and references will be given where my own attempts reach their limits.

Enterprises like this may become more important in the long run, given that it is rather unclear what kind of *independent* intuitions we could bring to bear on complexity/simplicity issues. At the most general level, the theory of competence as a whole is at stake. Could it be (much) simpler? This matter falls outside linguistics proper and into the realm of philosophy of science.[9]

> "The simplicity of linguistic theory is a notion to be analyzed in the general study of philosophy of science; the simplicity of grammars is a notion defined within linguistic theory" (Chomsky 1975/1955[1], p.119).

Internally, Chomsky (1995a) approaches Q1 from two sides. On the one hand, the computational system is constructed from elementary steps, "first principles," as it were, making "minimal" assumptions, reminiscient of the set-theoretic construction of natural numbers. On the other hand, the ultimate output of the system (and potential candidates of grammatical modules like Θ-Theory, Case Theory, Binding Theory, and the ECP) provide, sometimes implicit motivations for the direction and form theory construction is given. I would like to illustrate this constructive strategy in (12).

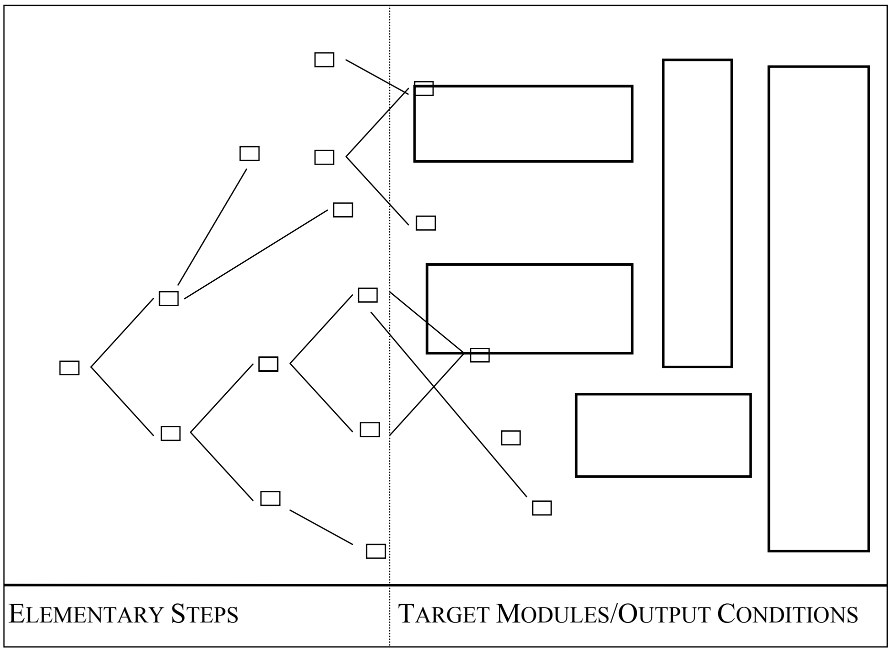
---

7 To develop a "really mechanical grammar" was considered important by Chomsky (1975/1955[1], p.69).

8 The importance of formalization itself has been controversial (cf. Chomsky 1990, Pullum 1989, and Fauconnier 1994).

9 For an early and non-trivial treatment of simplicity of theories see Goodman (1977/1951[1]).

(12)



Simplistic though it may be, this picture helps me to emphasize two things. First, it isn't clear where the dividing line between the two heuristics should lie. Yet, one can learn a lot about minimalism by watching out for it.

Secondly, it isn't clear whether one side couldn't render the other redundant. Pending radically new insight, however, this is a matter of faith. Do we mistrust the linguistic generalizations arrived at earlier? Then we might want to go into first principles much like Frege and Russell did when they tackled arithmetic at the turn of this century.[10] Or do we shy away from first principles, given that even the most innocuous ones force upon us quite specific further steps.

Indeed, it will turn out later on that neither side should be taken too seriously. There is reason to believe that certain of the initial steps taken on the grounds of "virtual

[10] Our faith in simple first steps might be propped up somewhat by reminding ourselves that it is cognitive science we are ultimately committed to: "Cognitive scientists (Kugler, Turvey, and Shaw (1980)) have pointed out the *first-order isomorphism fallacy* (FOIF), which consists in attributing to an organism internal structures analogous to the external structures of its outputs. Kugler, Turvey, and Shaw give many examples. For instance, termites build architecturally complex arches and pillars; the FOIF would consist in endowing the termites with some "mental program" for building arches and pillars. But in fact the termites "obey" a very simple rule in depositing glutinous sand flavored with pheromone: they follow a path of increasing pheromone density and deposit when the density gradient inverts. The fact that this behavior leads to the formation of arches and pillars is part of physics, not a property of termites" (Fauconnier 1994, p.168fn.3).

conceptual necessity," should be modified and enriched.[11] Likewise, I think that the actual impact of output conditions on minimalist syntax is rather weak and that more specific modules of grammar, although they provide the guidelines for research, can be reformulated quite freely. Let me formulate the overall intuition behind this in terms of the following metaphysical ("neo-romantic") hypothesis.

Hypothesis 1: The whole is more than the sum of its parts (=H1)

H1 will not be argued for explicitly later.[12] If anything, it is meant to draw attention to the non-trivial "summation" parts of syntax that tend to get blurred by atomist/lexicalist tendencies in minimalism.[13]

---

[11] Indeed, Chomsky (1998, p.122) adds the following caveat: "Note that the locution "optimal design" should always be read: *an* optimal design. What counts as optimal design might vary, depending on how the recursive procedures for forming LF structures are construed (even putting aside unclarities in the notion itself). The notion is relative to decisions about the nature of the language faculty that reach well beyond current understanding."

[12] See, however, section 2.3, for a brief comment.

[13] To end on a more sober note here, let me suggest a more literal version of the computational competence theory. That would mean taking complexity theory as defined in computer science to be a reasonable approach to Q1. It follows immediately that *time*, *hardware*, and *memory* are the resources expenditure of which is criterial for calling systems more or less simple. This is one of the things that, unfortunately, I won't have much to say about later. See for example Abeillé (1993), Kracht (1995, 1999, 2001), Rogers (1998), Stabler (1996, 1998), and Michaelis (2001), as well as papers in Kolb&Mönnich (eds.)(1999), for getting a clearer picture of what such speculations could be based on.

2 Minimalist Syntax

from swerve of shore to bend of bay
(James Joyce, Finnegans Wake)

The following section discusses “concepts and consequences“ of the minimalist program, as mainly introduced in Chomsky (1995a).[14] Section 2.1 recapitulates the “goals of linguistic theory,“ which among other things give rise to the distinction between “E-language“ and “I-language“ in generative research (cf. Chomsky 1986a). Section 2.2 discusses the interface-oriented construal of well-formedness in minimalist syntax, expressed in the “Principle of Full Interpretation.“ It is suggested that a fairly “narrow“ syntax-internal notion of interpretation may be sufficient for minimalist syntax (cf. 2.6.4, 2.7.1). Background information on the notions of “structure“ and “level“ is provided in section 2.3, in order to pave the way for an understanding of what giving up D-structure and S-structure, as discussed in section 2.4, implies for syntactic theory. The remainder of section 2.4 is devoted to the mechanism of feature checking, driving minimalist derivations. Also the question is raised whether the theta criterion of GB theory should continue to determine well-formedness in minimalism.

Sections 2.5 and 2.6 analyze two stages of reorganizing and unifying the phrase structural and transformational syntactic components on minimalist principles. This crucially involves (re-)definition of *binary* or “generalized“ transformations (2.5.1), later called “Merge“ (2.6.2), and *singulary* (“movement“) transformations (2.5.1), later called “Move“ (2.6.2). One major change concerns the free interspersal of these operations, another one the elimination of classical X-bar theory (2.5.1) in favor of contextually defined “bare phrase structure“ (2.6.4). This is accompanied by a turn from graph-based to set-based objects, auxiliary devices like nodes, bar-levels, and indices being (by and large) dispensed with (2.6.1). Sections 2.5.2 and 2.6.2 address the notion of “cyclicity,“ according to which each syntactic operation creates a new root node. It is shown that this may be empirically desirable (2.5.2). However, generalizing cyclicity by integrating Merge as a subroutine into Move leads to complications concerning the inventory of objects syntactic operations apply to (2.6.2). Section 2.6.3 discusses further complications of the same kind arising with the addition of labels and an adjunction operation to minimalist syntax. In section 2.6.4, the effect of interpretability on accessibility to syntactic operations, i.e. “visibility,“ is scrutinized wrt intermediate projections.

---

[14] See the introduction above for a caveat wrt Chomsky (2000, 2001).

Section 2.7, then, deals with the addition of “chains“ to the system, which results in conceptual as well as technical difficulties. These range from the unclear status of identity of “copies“ and interpretability of chains (2.7.1) to unwelcome multiplication of “checking resources“ (2.7.2). In section 2.7.3, a graph-theoretic solution to these problems is outlined, involving “multiconstituency“ or “multidominance.“ Section 2.7.4 addresses the viability of modifications made in Chomsky (2000, 2001) to deal with the same challenges.

Section 2.8, finally, provides a brief and partial look at the role of remaining (restricted) economy principles in minimalist syntax.

2.1 Goals of Linguistic Theory

A generative linguistic theory must, according to one of the still most succinct formulations (cf. Chomsky 1965, p.31), provide for

(1) a. an enumeration of the class $s_1, s_2, \ldots$ of possible sentences
b. an enumeration of the class $SD_1, SD_2, \ldots$ of possible structural descriptions
c. an enumeration of the class $G_1, G_2, \ldots$ of possible generative grammars
d. specification of a function *f* such that $SD_{f(i,j)}$ is the structural description assigned to sentence s_i by grammar G_j, for arbitrary i, j

By means of (1), generative linguists define the abstract linguistic competence underlying the remarkable ability of natural language users to (effectively) characterize natural language expressions as acceptable or unacceptable as well as to acquire and use a system of such expressions.

Given that at least the objects of (1a) and (1b) constitute infinite sets, different ones for different languages at that, a lot of ingenuity has to go into a finite characterization of the tools for enumeration, recursion being among the most valuable ones. By and large, one can say that work on (1a)/(1b) has kept generative linguists busy ever since the goal was thus defined, while (1c)/(1d) has, for various reasons, been much less focused on. This study will not be an exception in that respect.

Now, on the assumption that a particular language *L* includes a fixed (ideally finite) vocabulary *A*, the set to be enumerated according to (1a) can be seen as a proper subset of A^*, the Kleene closure of *A*. This set, call it L^A, has sometimes been called an “E-language“ (cf. Chomsky 1986a), because it relates to among other things the *external* domain of “observable“ raw data (linguistic corpora etc.). What is involved in “generating“ such an E-language has been explored by mathematicians as well as formal linguists. Early results from that domain strongly influenced the foundations of generative linguistics, as can be seen from Chomsky (1975/1955[1], 1965) and its use of concatenation algebras and context-free string rewriting systems.[15]

[15] For some history of generative grammar, see for example Heny (1979), van Riemsdijk&Williams (1986), as well as articles in Jacobs et al. (eds.) (1993).

Nevertheless, Chomsky has repeatedly argued that linguistically relevant facts concern the higher-order structure associated with the individual strings (= “sentences“) in L^A. Insightful generalizations about the language user seem to be derivable from the way sentences are organized underlyingly (e.g. being broken up into several layers of constituents) much more directly than from the shape of possible subsets of A*. The following quote from Chomsky (1963, p.326) makes this point in terms of the two kinds of complexity measure distinguished in formal linguistics. These measures impose a ranking on grammars. Thus, “*weak* generative capacity“ relates to grammars seen as defining subsets of A*, i.e. E-languages, while “strong generative capacity“ relates to the kind of structures grammars can assign to the strings of L^A.

> “Ultimately, of course, we are interested in studying strong generative capacity of empirically validated theories rather than weak generative capacity of theories which are at best suggestive. It is important not to allow the technical feasibility for mathematical study to blur the issue of linguistic significance and empirical justification. We want to narrow the gap between the models that are accessible to mathematical investigation and those that are validated by confrontation with empirical data, but it is crucial to be aware of the existence and character of the gap that still exists. Thus, in particular, it would be a gross error to suppose that the richness and complexity of the devices available in a particular theory of generative grammar can be measured by the weak generative capacity of this theory. In fact, it may well be true that the correct theory of generative grammar will permit generation of a very wide class of languages but only a very narrow class of systems of structural descriptions, that is to say, that it will have a broad weak generative capacity but a narrow strong generative capacity.“

Indeed, from fairly early on, attention has been diverted away from E-languages, and complexity issues got discussed to a much lesser degree. Empirical validation of models, it is fair to say, is still a central issue, while narrowing the gap to mathematical investigation has not been so high on the agenda, at least as far as the Chomskyan branch of generative linguistics is concerned. It will become clear later on that minimalism, as already pointed out in section 1.1, would seem to benefit from narrowing the gap in order to be a more transparent enterprise (cf. Kolb 1997a).[16]

Generally, the structural properties to be captured under (1b) have been considered to arise from the internalized knowledge of language attributed to language users as an innate endowment. Thus, in keeping with a concern for the “internal“ states of affairs cognitive science occupies itself with and heeding Chomsky's (1963) warnings, a majority of generative linguists has concentrated on exploring the proper formulation of (1b). The machinery supposed to do the required enumeration has consequently been called “I-language“ (Chomsky 1986a).

One obvious question arises at this point. How can the implied universalist view be compatible with the diversity of human languages we are familiar with? This is

[16] For recent discussion of strong generative capacity, see Joshi (1985), Kracht (1995), Stabler (1996), Rogers (1998), and Michaelis (2001), as well as papers in Kolb&Mönnich (eds.) (1999).

immediately relevant to (1c) and therefore (1d). The set of possible grammars must be broad enough to accommodate that diversity. At the same time it must be narrow enough for the right grammar to be discoverable in the short time allotted to children for language acquisition, given only "positive" input data.[17] Thus, issues of finiteness may arise here (cf. Chomsky 1981, Pullum 1983).

Early on, it was proposed that grammars can be further ranked according to a simplicity measure, "a function of the number of symbols in the grammar" (Chomsky 1975/1955[1], p.117). While viewed by critics as "vague talk,"[18] the proposal was recognized immediately as being in need of the empirical discovery of generalizations. These were to provide formal and substantive universals, which could realistically constitute I-language and narrow the choices to be made under language acquisition.

> "The major problem in constructing an evaluation measure for grammars is that of determining which generalizations about a language are significant ones" (Chomsky 1965, p.42).

Driven by this task, generative grammar has undergone various changes,[19] culminating in the so-called "principles and parameters" approach.[20] Under this approach, a set of highly abstract universal principles defining a number of grammar "modules" is considered to be combined with a finite number of parameters, the fixing of which, ideally a binary choice, is all there is to language acquisition. For example, the child has to discover about her language whether or not lexical heads like verbs and prepositons precede their complements, i.e. she has to set the "head parameter."[21]

Not surprisingly, none of the parameters proposed so far has remained unchallenged. Thus, currently we see (at least) two major strands within generative syntax aimed at reducing possible language variation even further. Kayne (1994) considers the phrase structure of natural language to follow a uniform pattern of basic rightward branching.[22] Accordingly, heads always precede their complements by universal hypothesis, which would imply that no head parameter has to be set. At the same time, inspired among others by the works of Borer (1984), Pollock (1989), Ouhalla (1991, 1992), and

---

[17] No corrections and no indication that some expression does not belong to L^A, i.e. no "negative" input data, are considered to be available to the child.

[18] "Until we understand the strategies which make general learning possible - and vague talk of 'classes of hypotheses' and 'weighting functions' is utterly useless here - no discussion of the *limits* of learning can even begin" (Putnam 1975, p.116). See also Peters (1972) for more constructive criticism.

[19] See for example Chomsky (1971), Jackendoff (1972), Chomsky (1977). See Leuninger (1979) for an overview over this intermediate period.

[20] Chomsky (1981, 1982, 1995a, chapter1 [= with Howard Lasnik]).

[21] For some recent formalizations of such a system, see Gibson&Wexler (1994), Berwick (1994), Berwick&Niyogi (1996), and Frank&Kapur (1996), the latter forcefully reminding us that finiteness isn't the only criterion in that area, given that only 20 binary parameters define a class of over one million grammars. For a critical discussion of parameters and an alternative based on the earlier "length of grammar" approach, see Fodor&Crain (1990).

[22] See Haider (1992) for a closely related proposal.

Guilfoyle&Noonan (1988), Chomsky (1995a) assumes as one of the tenets of the minimalist program that parametric choices are limited to properties of functional heads, determiners, complementizers, and abstract auxiliary-like elements among them. More narrowly, the task of acquiring syntax consists to a large extent in discovering the (binary) "strength" value of (features of) such functional heads. The "strength" property is hypothesized to be responsible for word-order variation among natural languages.[23] If true, learning the "rules" of grammar would be aligned with acquiring a lexicon, i.e. a subset of A, the need for the latter being one of the most uncontroversial views among linguists.[24] Consequently, if both attempts at reduction just sketched succeed, task (1c) would be less worrisome. Of course, it remains to be seen whether the structure of the lexicon as well as (universal) morphology and phonology can be dealt with as elegantly.[25]

Furthermore, one has to bear in mind that generative grammars tend to have to idealize quite considerably. Indeed, principles and parameters define only a "core" subset of each L^A. Marked extensions are relegated to a "periphery" of grammar, incorporating language particular rules of various provenance.[26] The reader should thus not be surprised at the fact that in the following, a fairly narrow set of sentence structures will undergo what may seem to be overly microscopic examination.

2.2 Interfaces and Well-Formedness

Let's now turn to the way the "computational system of human language," C_{HL}, i.e. the most recent version of I-language (cf. section 1.1 above), can derive infinite sets of structural descriptions in accordance with (1b) of section 2.1. This is where centrally important answers to the minimalist "Gretchenfrage" Q1 come to the fore.

Q1: How "perfect" is language?

Given that the rules and representations of grammar must minimally relate sound and meaning, they won't, in the minimal of all possible worlds, do more than that. Thus, according to Chomsky (1995a, p.219, p.225), C_{HL} interfaces with the "conceptual-intentional" (C-I) as well as the "articulatory-perceptual" (A-P) components of the mind/brain at exactly two points, an LF-representation λ and a PF-representation π.

---

23 See Lasnik (1999). Chomsky (2000, 2001) reanalyzes strength as the presence (vs. absence) of an abstract feature triggering movement.

24 Manfred Bierwisch (p.c.) points out that the complexity of acquiring the lexicon depends on how rich the structure of the latter is taken to be. This, of course, is a highly controversial matter.

25 For the latter issues, see among many others Bobaljik (1995b), Halle&Marantz (1993), Kenstowicz (1994), and Williams (1989).

26 Cf. Chomsky (1981). See Fodor&Crain (1990), arguing for an alternative theory, which they claim exhibits a more natural continuity between core and peripheral constructions of language.

(2) “C_{HL} generates linguistic expressions L, which consists of pairs $<\pi,\lambda>$, π a PF-representation and λ an LF-representation.“

The sets of structural descriptions of (1b) will therefore contain $<\pi,\lambda>$-tuples. PF in (2) refers to the familiar part of grammar called “phonetic form,“ while LF is a shorthand for “logical form,“ the latter defining an, often invoked, but seldom fully specified, logico-syntactic structure that will undergo semantic translation. Unfortunately, the exact nature of LF is subject to massive controversy,[27] which, as we are going to see, negatively affects the intelligibility of the claim that the minimalist enterprise constitutes more than a rival theory of *syntax*.[28]

---

[27] See for example contributions to *Linguistics and Philosophy 12* (1989).

[28] For a number of reasons, I'm being vague at this point. First of all, the exact nature of PF and LF cannot be appreciated before the technical notion of a linguistic “level“ has been introduced and discussed (see 2.3 below). In fact, PF and LF are called the interface “levels.“ So, jumping ahead somewhat, it looks as if all of syntax is included in LF while all of morphology and phonology belongs to PF. From this it can be concluded, secondly, that usage of the term “the interface“ is itself slightly vague. Thus, compare the following quotes from Chomsky (1995a):

(A) “Conditions on representations - those of binding theory, Case theory, Θ-theory, and so on - hold only *at the interface*, and are motivated by properties *of the interface*, perhaps properly understood as modes of interpretation by performance systems“ (p.170f; italics mine, H.M.G.).

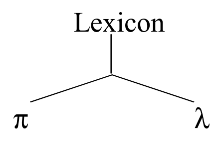
(B) “Notice that I am sweeping under the rug questions of considerable significance, notably, questions about what in the earlier Extended Standard Theory (EST) framework were called “surface effects“ on interpretation. These are manifold, involving topic-focus and theme-rheme structures, figure-ground properties, effects of adjacency and linearity, and many others. Prima facie, they seem to involve some additional level or levels internal to the phonological component, postmorphology but prephonetic, accessed *at the interface* along with PF (Phonetic Form) and LF (Logical Form)“ (p.220; italics mine, H.M.G.).

In order to make sense of (A), assuming standard versions of binding theory etc., “the interface“ has to refer to the designated LF-representation λ. At the same time, (B) singularizes “the interface“ in a way that, if I understand this correctly, hypothetically unifies the interface levels LF, PF, and the putative additional one, into a superstructure, i.e. “the interface,“ accessible via a set of representations, among which, of course, π and λ would have to be found.

Thirdly, I prefer not to go into what I've called “semantic translation.“ Whether or not semantics is inside C_{HL} would be more than a terminological question if we knew more precisely what conceptual-intentional (C-I) systems are supposed to be. I am, however, inclined to believe that there is something genuinely linguistic about matters of semantic interpretation, which would qualify them as “knowledge of language,“ the object of study of generative linguistics. Natural language syntax as it emerges from studies of generative syntacticians does not seem to me to encompass these matters of interpretation in any sufficient way. For detailed discussion of some of these questions see for example Bierwisch (1982), Chomsky (1971, 1995c), Higginbotham (1985), Jackendoff (1972, 1990), Partee (1975), and von Stechow (1993). Remarks like the following terminologically complicate the picture further, given that the term “syntax“ would seem to be construable as covering large parts of cognitive science. My general point concerning interpretation, though, is unaffected if not vindicated here. “[. . .] I suspect that much of the very fruitful inquiry and debate over what is called “the semantics of natural language“ will come to be understood as really about the properties of a certain level of syntactic representation - call it LF - which has properties developed in model-theoretic semantics, or the theory of LF-movement, or

Schematically, C_{HL} will look like (3), a mapping between three objects, π and λ having been introduced above already. The "lexicon" takes over the role of vocabulary A of section 2.1 in a non-trivial way.

(3)



It is the job of the generative linguist to determine the operations ("rules") and data structures ("representations") underlying (3). The picture will be gradually filled in below. Most importantly, some kind of well-formedness condition has to be imposed on (2) and (3), such as the one given in (4).

(4) A linguistic expression $<\pi,\lambda>$ is *grammatical* iff there is a derivation D that yields $<\pi,\lambda>$, and π and λ each satisfy the "Principle of Full Interpretation" (FI).

Indeed, emphasizing the essentially procedural character of C_{HL}, Chomsky (1995a, p.219f) gives a slightly different formulation of well-formedness.

(5) "A derivation *converges at* one of the interface levels if it yields a representation satisfying FI at that level, and *converges* if it converges at both interface levels, PF and LF; otherwise it *crashes*.

Highlighting derivations over representations becomes significant as soon as one recognizes the *two-layered* approach to well-formedness, now called "convergence."

> "The language L thus generates three relevant sets of computations: the set D of derivations, a subset D_C of convergent derivations of D and a subset D_A of admissible derivations of D. FI determines D_C, and the economy conditions select D_A. [. . .] D_A is a subset of D_C" (Chomsky 1995a, p.220).

The structural descriptions C_{HL} designates as well-formed are the ones that (i) satisfy FI and (ii) are defined by a derivation $D \in D_A$.

> "It seems that a linguistic expression of L cannot be defined just as a pair (π,λ) formed by a convergent derivation. Rather, its derivation must be *optimal*, satisfying certain natural economy conditions [. . .]" (Chomsky, 1995a, p.220).

---

something else, but which belongs to syntax broadly understood - that is, to the study of mental representations and computations - and however suggestive it may be, still leaves untouched the relations of language to some external reality or to other systems of the mind" (Chomsky 1991, p.38).

Although I will deal with economy conditions only much later (see section 2.8), this is the place to point out that the version of minimalist syntax I'm going to advocate will not contain economy conditions. Thus, I consider (4), slightly modified later, to be an adequate well-formedness condition for C_{HL}.

What we still don't know about (4) is, first of all, what a derivation should look like. This difficult issue will be dealt with at length below. Secondly, we have to understand what FI corresponds to. This is made explicit below (cf. Chomsky 1995a, p.27).

(6) The Principle of Full Interpretation (FI)
A representation R *satisfies FI* iff R does not contain uninterpretable ("superfluous") elements.

FI reflects one of the most basic principles of minimalism. Elements of C_{HL} are taken to provide "instructions" to the C-I and A-P systems of the mind/brain. Ideally, therefore, elements that do not do this, i.e. elements that are not "interpretable" from the outside in this sense, should not occur at the interfaces. We are going to see that ample use is made of this heuristic guideline when motivating the inner workings of derivations. Thus, a need to eliminate or neutralize uninterpretable elements functions as a *causa efficiens*. Under this perspective, C_{HL} observes some kind of "economy of representations."[29]

We are also going to see later that extensions of FI are more problematic than might be expected at first sight. Thus, the typology of "legitimate LF-objects" satisfying FI (cf. Chomsky 1995a, p.153f), though arguably desirable, forces a sharpening of notions in the domain of chains (cf. section 2.7). The general difficulty has to do with so-called "visibility" of syntactic elements to syntactic operations, a notion suggested to be derivable from FI. I will criticize this notion in section 2.6.4. If that argumentation is correct, it appears that where interpretability is appealed to *within* C_{HL}, this notion is unlikely to have much to do with interpretability as definable from the *outside* (C-I/A-P systems). It follows that unless genuinely semantic (or sensorimotor) principles are brought to bear on C_{HL}, the term "interpretability" is rather misleading and perhaps avoidable.[30] More concretely, FI reduces to a narrow syntax-internal condition on well-formedness (cf. section 3.1).

Another ingredient in the abstract design of C_{HL} is furnished by the following uniformity principle (cf. Chomsky 1995a, p.229).

---

[29] Cf. Chomsky (1995a, chapter 2). More explicitly "[t]he linguistic expressions are the optimal realizations of the interface conditions, where "optimality" is determined by the economy conditions of UG" (Chomsky 1995a, p.171). I hesitate to consider FI an "economy principle," as will become apparent from the discussion in section 2.8.

[30] Note that *there*-constructions, which provided one of the major empirical arguments for FI-driven operations, i.e. LF-replacement (cf. Chomsky 1986a), has seen a number of syntactic revisions such that actual replacement is no longer assumed (cf. Lasnik 1995, 1999a). On the original view, thus, *there* would strictly speaking violate FI at LF. However, if recent semantic approaches to these constructions are on the right track (cf. Blutner 1993 and Musan 1996), this may not be so. For further discussion see also Kempson (1988).

(7) Computational procedures are uniform

This means that the same kinds of operation generate π as well as λ. However, a caveat is added immediately. (7) is supposed to hold

> "[. . .] not for the computation from N to π; the latter modifies structures (including the internal structure of lexical entries) by processes very different from those that take place in the N→λ computations" (Chomsky 1995a, p.229).

N, henceforth notated as ℕ, is an intermediate data structure called "numeration" to be described later (cf. section 2.6.1). We can conclude that the branch toward π in (3) above is of a different nature from the one leading to λ. The former, of course, is usually called morphophonology and well-known for its recalcitrant puzzles.

If, however, (7) doesn't hold across the board, i.e. if uniformity of operations is not a defining property of C_{HL}, the system is not unlikely to regroup into a syntactic and a morphophonological component, the latter often referred to as "PF-component." Given that, it is not unreasonable to follow Jackendoff (1997) in considering the minimalist program as essentially a program for ("narrow") *syntax*,[31] embedded in the not unusual tripartite structure in (8).

(8) PHON - SYN - SEM

Interfacing with (C-I) and (A-P) systems would occur at the obvious places. If accurate, this means that the minimalist program has to be reconstructed as *minimalist syntax*, a framework to be compared with earlier models such as GB and rival approaches such as HPSG, LFG, and TAG (cf. Abeillé 1993 and Horrocks 1987). Symptomatically, Chomsky (1995a) contains practically no considerations of any semantic or (genuinely) morphophonological nature. I will therefore, right at this outset, suggest a modification of the well-formedness conditions for C_{HL} in terms of (9) and (10).[32]

(9) C_{HL} generates *linguistic expressions L*, which consist of triples
<*sem*,*syn*,*phon*>, *phon* a PF-representation, *syn* a syntactic representation, and *sem* a semantic representation.

(10) A linguistic expression L (= <sem,syn,phon>) is *grammatical* iff
a. there is a derivation D that generates L, and
b. *sem* and *phon* are well-formed, and
c. *syn* can be translated into *sem* and *phon*

The rest of this study can be understood as concerning the nature of *syn*-representations as well as the derivations generating them. Together, these determine the SYN-

[31] Cf. Chomsky (2000, p.100).
[32] See also section 3.1, below.

component of C_{HL}. The notion of “translatability“ introduced in section 3.1, is simply the “narrow“ variant of FI.[33]

Of course, it has to be kept in mind that the minimalist program is intended to open a much broader perspective. This, however, mainly affects style of argumentation but doesn't seem to me to really matter when it comes to concrete syntactic detail. As can be inferred from the preceding chapter, I'm advocating the project of “narrow syntax“ (cf. Chomsky 1995a, p.34).[34]

2.3 Structure and Levels

By way of filling in the picture charted in (3) of section 2.2, I will go over some fairly basic aspects of syntactic theory, which, I hope, is going to unclutter later discussion. For obvious reasons, “structure“ is a core concept of generative linguistics in general, and generative syntax in particular.[35] However, if asked to explain that term, I find myself in the kind of predicament St. Augustine describes when he confronts the concept of time: “If nobody asks, I know it, if I try to explain it, I don't (know it).“[36] Indeed, I've referred to “structure“ a couple of times already, relying on intuition.

Now, it is generally taken for granted that linguistic expressions are *structured* entities. When trying to “break them up,“[37] we are likely to invoke units such as sentences, words, and syllables. Indeed, thinking about language in the abstract easily draws one into some kind of patterning activity.[38] (Linguistic) “Structuralism“ can be said to have identified (important parts of) language with its structure, thereby directing attention to the linear “arrangement“ of linguistic units, the “syntagmatic“ aspect of language, and the mutual exchangeability of specific units of language, its

---

33 Note that such a formulation of well-formedness leaves a number of options. If it is decided that e.g. binding conditions affect “grammaticality,“ violations can be encoded in a way that bar translation into either *phon* or *sem* or both. The current state of minimalist syntax doesn't allow any solid conclusions here. I will provide some discussion of Θ-theory shortly (section 2.4).

34 Whether anything like the “end of syntax“ (Marantz, 1995) is implied by either of the above alternatives is doubtful, if not a “category mistake“ (cf. section 2.3).

35 Cf. expressions like “the structure of language,“ “syntactic structures,“ etc.

36 “Si nemo a me quaerat, scio, si quaerenti explicare velim, nescio.“ Quoted from E.Husserl (1980/ 1928[1], p.2).

37 Clearly, this kind of terminology is unavoidable. What I have to say in this chapter does, of course, affect terms like “unit,“ “order,“ and “form“ as well.

38 This intuitive familiarity with pattern can explain the air of triviality evoked by such sweeping statements as “Das Erscheinungsbild der Wirklichkeit ist stark strukturiert“ (“Reality appears to be highly structured.“) Yet this is how Manfred Eigen and Ruthild Winkler (1985, chapter 6) open their exciting and nontrivial discussion of what games, physico-chemical patterns, and genetics have in common.

“paradigmatic“ aspect.[39] To put it differently, language is realized physically, i.e. spatio-temporally (sequences of marks on paper, articulatory gestures, sound waves etc.)[40] and its units reflect various layers of categorization.

Now, given that structuralism paved the way for full-fledged theory construction, we have to look more closely at the “identification of language with its structure.“ Pointing out the syntagmatic and paradigmatic aspects of language as such remains in an intuitive pretheoretic area. In that area, the term “structure“ most clearly links up with it's “metaphorological“ origin in the domain of “building.“[41] An axiomatic theory of language has to put this pretheoretic background on a well-defined basis. Chomsky (1975/1955[1]) solves this problem by investing “concatenation,“ +,[42] into the theory, a binary, associative, noncommutative, algebraic operation for constructing complex linguistic expressions from atomic ones at various levels. In such a system, the syntagmatic aspect of language will be *there* by definition.[43] Subtle, though this point may seem at this early stage, we can already say that expressions, called “strings,“ like (11a)/(11b) are well-defined, while (11c)/(11d) is not defined in the system, no matter how much the latter may appear to be “structured“ *prima facie*. More precisely, although each individual word is considered a string (of length one), the linear arrangement of words constitutes a string only in (11a) and (11b), not, however, in (11c) and (11d). (Take *a*, *b*, and *c* to be words for the sake of argument.)

(11) a. a + b + c
b. The + borogoves + greet + the + toves
c. a b c
d. The borogoves greet the toves

Most linguistic texts leave out the concatenation operator and rely on the physical arrangement as well as the charity of the reader when it comes to filling in such seemingly trivial detail. I will stick to that convention and use + only where its presence is under debate or misunderstanding has to be avoided.

Note that this rather effortless first step could probably not have been made without the analytical labor that went into laying the foundations of mathematics at the turn of the (last) century.

---

39 Cf. de Saussure (1969/1915[1], chapter 5). I'm simplifying somewhat. De Saussure uses the broader term “rapports associatifs“ for the paradigmatic aspect. Language as a “system of oppositions“ is the most comprehensive notion Genevan structuralism is famous for (cf. Jakobson 1974/1939[1]).

40 Chomsky (1981, p.10) briefly discusses the matter of epistemological priority, likely to guide the search for primitives of linguistic theory. Thus, it would “be reasonable to suppose that such concepts as “precedes“ or “is voiced“ enter into the primitive basis [. . .].“

41 Cf. Kambartel (1968, p.181) and Gordon (1978).

42 This notation is taken from Chomsky (1957).

43 More abstract solutions based on different operations are possible as we are going to see in later sections (cf. Curry 1961). Even concatenation was viewed as a more abstract device by Chomsky (1975/1955[1], p.380), not necessarily related to “temporal order of sounds.“

> "Thus what we defined as the "relation-number" is the very same thing as is obscurely intended by the word "structure" - a word which, important as it is, is never (so far as we know) defined in precise terms by those who use it" (Russell 1993/1919[1], p.61).

Russell defined "structure" in the same way he defined "number," the latter being considered a property of (classes of) classes, the former a property of (classes of) relations. So today, the most neutral and arguably most useful concept of structure derives from mathematics, which recognizes two kinds of structure. An "algebraic structure" is a set on which at least one relation or at least one operation is defined, a "topological structure" is a set with a designated system of subsets (cf. Knerr 1991, p.376).[44]
These definitions already enable us to go back and motivate hypothesis 1 of section 1.1.

H1: The whole is more than the sum of its parts

Thus, it is crucial to distinguish the exact kind of operation, relation or property that turns a set into a structured whole, i.e. into a "structure." Failure to do so can complicate linguistic discussions considerably, as we are going to see in later sections when we have to clarify the interplay of features, lexical items, and constituents.

Now, American structuralism approximated the mathematical perspective as the following remark by Harris (1981/1954[1], p.3) indicates.

> "For the purposes of the present discussion, the term structure will be used in the following non-rigorous sense: A set of phonemes or a set of data is structured in respect to some feature, to the extent that we can form in terms of that feature some organized system of statements which describes the members of the set and their interrelations (at least up to some limit of complexity)."

A further step into a theory of syntactic structure was Chomsky's (1975/1955[1]) formalization of "immediate constituent analysis" (ICA) in terms of "phrase markers." ICA was used by American structuralists to capture the hierarchical organization of strings of words into larger units up to sentence level. This is where the well-known diagrams of generative syntax, usually called "phrase markers" come from.[45] Thus, (11b) could be represented as (12a) or (12b), the latter being equivalent notations of the phrase marker associated with (11b).

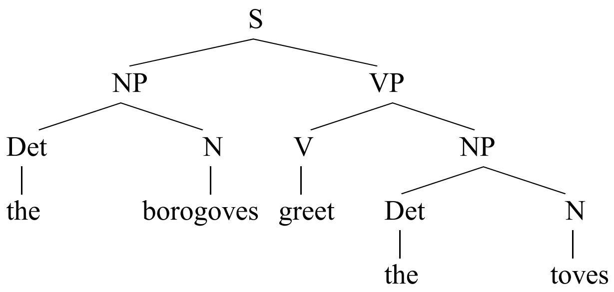
---

44 As a consequence of mathematical axiomatization, structure assumed an important role in one of the most delicate areas of the philosophy of science, namely, the question of implicit definitions (cf. Essler 1982, Kambartel 1968, and Stegmüller 1987).

45 There is an extra subtlety involved here, since Chomsky didn't fully reconstruct ICA. Therefore, certain structures recognizable in ICA were banned from phrase markers. See Chomsky (1975/1955[1] chapter 7), Manaster-Ramer&Kac (1990), and Blevins (1990) for discussion.

(12) a. [$_{S}$ [$_{NP}$ [$_{Det}$ The] [$_{N}$ borogoves]] [$_{VP}$ [$_{V}$ greet] [$_{NP}$ [$_{Det}$ the] [$_{N}$ toves]]]]

b.



For quite obvious reasons, the lexical items are called "terminals" whereas the other categories are called "nonterminals."

At least within generative linguistics, being able to put the sentences of natural language into this format continues to be both a heuristic for the practical analysis of sentences and a well-formedness criterion imposed on the structural descriptions to be enumerated by the computational system.[46]

Now, what (12) gives us is a very simple way of looking at "constituents." Thus, any sub-phrase marker defines a constituent of the phrase marker it is a part of. The NP built up from the terminals *the borogoves* is an immediate constituent of S and the one built up from *the toves* is an immediate constituent of VP and a constituent of S. The category labels encode part of the paradigmatic aspect of language, that is, they correspond to "distributional properties" in the terminology of American structuralism. Thus, to the extent that elements can be substituted for each other, they form a class or "category" which can be assigned a uniform label, as illustrated in (13).[47]

(13) a. [$_{NP}$ Humpty Dumpty] greets the toves
b. [$_{NP}$ The Queen of Hearts] greets the toves
c. [$_{NP}$ She] greets the toves

Formally, phrase markers like (12) can be defined as "constituent structure trees" (Partee et al. 1993, p.441f).[48]

---

[46] Of course, further conditions are necessary for linguistic theory to be restrictive, given that "[v]irtually anything can be expressed as a phrase marker, i.e. a properly parenthesized expression with parenthesized segments assigned to categories" (Chomsky 1972, p.67).

[47] Cf. section 2.6.3. When it comes to entire texts, not to speak of languages, the complexity of distributional analysis easily gets almost insurmountable, a fact which necessitated sophisticated techniques and ultimately appeal to hypothetico-deductive methods instead (cf. Harris 1981, Chomsky 1975/1955[1]).

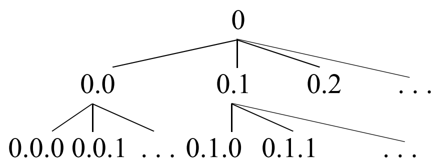
[48] Cf. McCawley (1968, p.244). Within graph-theory, nodes are called "vertices" and the ordered pairs of immediate successors in the dominance relation are called "edges." Where left-to-right order of edges matters, i.e. where a precedence relation is defined, trees are called "ordered trees" (Cf. Cormen et al. 1990, chapter 5).

(14) A *constituent structure tree* is a mathematical configuration $\langle N, Q, D, P, L \rangle$, where
N is a finite set, the set of nodes
Q is a finite set, the set of labels
D is a weak partial order in $N \times N$, the dominance relation
P is a strict partial order in $N \times N$, the precedence relation
L is a function from N into Q, the labeling function
and such that the following conditions hold:
(1) $(\exists x \in N)(\forall y \in N)[\ \langle x,y\rangle \in D\ ]$ (Single Root Condition)
(2) $(\forall x,y \in N)[\ (\ \langle x,y\rangle \in P \vee \langle y,x\rangle \in P\ ) \leftrightarrow (\ \langle x,y\rangle \notin D \wedge \langle y,x\rangle \notin D\ )\ ]$
(Exclusivity Condition)
(3) $(\forall w,x,y,z \in N)[\ (\ \langle w,x\rangle \in P \wedge \langle w,y\rangle \in D \wedge \langle x,z\rangle \in D\ ) \rightarrow \langle y,z\rangle \in P\ ]$
(Nontangling Condition)

With the help of (14), well-known concepts like C-Command can be defined. Conditions (1) - (3), though linguistically motivated, can be and have been modified to some extent as we are going to see later.[49]

Note that there is a subtlety involved in comparing (12) and (14). Order in trees is defined on *nodes*, not *labels*, given that identical labels can be assigned to different portions of the structure identified by the nodes that dominate them. This makes computation over and reasoning about trees nontrivial. One way to ease that problem is to introduce an addressing system based on numbers, where, to give just one example, nonfinal integers encode the address of the immediately dominating node and the final integer indicates how many other siblings precede a node below the immediately dominating node.[50]

(15)



Otherwise, more or less cumbersome definite descriptions like “the node labeled NP immediately dominated by the node labeled S“ have to be used. This kind of subtlety becomes quite crucial in the constructions of minimalist syntax to be discussed later.

Phrase markers like (12) can be converted into strings by a recursive top-down procedure applying to each tree such that either the root of that tree is listed or alternatively the roots of its subtrees are listed from left to right. Thus, the stringset yielded by (12) is in part indicated in (16).

---

[49] Cf. section 3.2. Condition (3) had been introduced by Chomsky (1975/1955[1], p.175) as an axiom of the phrase structure level, a departure from at least some versions of ICA as has been pointed out by Blevins (1990).

[50] Cf. the definition of “tree domains“ in Johnson (1988, p.77).

(16) { S, NP+VP, Det+N+VP, the+N+VP, . . . , NP+V+NP, NP+greet+NP, NP+V+Det+N, NP+V+the+N, NP+V+Det+toves, . . . , Det+N+V+Det+N, . . . , the+borogoves+greet+the+toves }

Objects like these constitute the original conception of phrase marker as formalized in Chomsky (1975/1955[1]). They mirror a view on grammar inspired by string rewriting systems that unified automata theory and axiomatic proof theory and gave rise to programming languages like “Algol.“ Thus, given a single “axiom“ S, plus a set of productions like the ones in (17)

(17) a. S $\rightarrow$ NP+VP d. Det $\rightarrow$ the g. V $\rightarrow$ greet
b. NP $\rightarrow$ Det+N e. N $\rightarrow$ borogoves
c. VP $\rightarrow$ V+NP f. N $\rightarrow$ toves

strings of terminals underlying natural language sentences could be “proved“ as “theorems“ of the system.

Wrapping up the discussion so far, we can define a “constituent structure grammar“ (or “phrase-structure grammar“) $\mathbb{G}$ as follows.[51]

(18) $\mathbb{G} = < V, +, \rightarrow, V_T, S, \# >$ is a system of concatenation meeting the following conditions:
1. V is a finite set of symbols called the *vocabulary*. The strings of symbols in this vocabulary are formed by concatenation; + is an associative and noncommutative binary operation on strings formed on the vocabulary V.
2. $V_T \subset V$. V_T we call the *terminal vocabulary*. The relative complement of V_T with respect to V we call the *nonterminal* or *auxiliary* vocabulary and designate it by V_N.
3. $\rightarrow$ is a finite, two-place, irreflexive and asymmetric relation defined on certain strings on V and read “is rewritten as.“ The pairs (φ,ψ) such that $\varphi \rightarrow \psi$ are called the (*grammatical*) *rules* of G.
4. Where $A \in V$, $A \in V_N$ if and only if there are strings φ, ψ, ω such that $\varphi A\psi \rightarrow \varphi\omega\psi$. $\# \in V_T$; $S \in V_N$; $e \in V_T$; where # is the *boundary symbol*, S is the *initial symbol* that can be read *sentence*, and *e* is the identity element with the property that for each string φ, $e\varphi = \varphi = \varphi e$.

It is quite uncontroversial that something like $\mathbb{G}$ should play an important part in a theory of natural language syntax.[52]

---

51 Slightly adapted from Chomsky&Miller (1963, p.292). I sidestep the more careful development of phrase structure in Chomsky (1975/1955[1]) and its myriad formal consequences (cf. Chomsky&Miller 1963 and Chomsky 1963). Note that the productions in $\rightarrow$ can also be seen as “node admissibility conditions,“ which differ from string rewriting at the algorithmic level (cf. McCawley 1968, Gazdar et al. 1985), or “tree formation rules“ (cf. Partee et al. 1993, Manaster-Ramer&Kac 1990). A more compact use of stringsets has been made by Lasnik&Kupin (1977), who formalized the then current version of transformational grammar.

Chomsky (1975/1955[1]) considered $\mathbb{G}$ to provide a major part of the "linguistic level" called "phrase structure." This is fully in line with the attempt to give a definition of "language" in terms of "linguistic levels" as defined below (Chomsky 1975/1955[1], p.108).[53]

(19) A *level* $\mathbb{L}$ is a system $\mathbb{L} = \langle L, +, R_1, \ldots, R_m, \mu, \Phi, \varphi_1, \ldots, \varphi_n \rangle$, where
(i) $\mathbb{L}$ is a concatenation algebra with L its set of primes
(ii) $R_1, \ldots, R_m$ are classes and relations defined within $\mathbb{L}$. R_1 is the identity relation =
(iii) μ is a set of $\mathbb{L}$-markers - elements of some sort constructed in $\mathbb{L}$
(iv) Φ is a mapping which, in particular, maps μ into the set of grammatical utterances
(v) $\varphi_1, \ldots, \varphi_n$ express the relations between $\mathbb{L}$ and other levels

"Language," in the technical sense of I-language, could then be defined as a (hierarchical) sequence of levels, illustrated in (20).

(20) *Language* = $\langle \mathbb{Pn}, \mathbb{Pm}, \mathbb{C}, \mathbb{W}, \mathbb{M}, \mathbb{P}, \mathbb{T} \rangle$

Included are the levels of "phones," "phonemes," "syntactic categories," "words," "morphemes," "phrase structure," and "transformations," respectively.[54] Fleshed out on the basis of results obtained in the respective fields of research, (20) would provide the appropriate structures for meeting the "goals of linguistics" sketched in section 2.1.

Indeed, something like (19) is still assumed to unify generative linguistics (Chomsky 1986a, p.46, Reuland 1988).[55] Yet the sheer amount of modifications, due to empirical insight and theoretical reorganization, has not been accompanied with a careful development of the appropriate notions, while terminological additions like "component," "stratum," and "system" further complicate that task.[56] Anyway, it is exactly in this area of higher-order constructs that disagreement prevails, not the least because assumptions are massively underdetermined by empirical facts. So, it appears wise not to take arguments based on these concepts too seriously unless worked out in detail.

---

52 Controversy arises wrt the questions what additional machinery is needed and whether $\mathbb{G}$ can be unified with that machinery. See for example Abeillé (1993), Baltin&Kroch (eds.)(1989), Brody (1995), Chomsky (1995a), Gazdar et al. (1985), Horrocks (1987), Joshi (1985), as well as Pollard&Sag (1987, 1994).

53 $\mathbb{L}$ also contains an identity element (p.106) and to "simplify the constructional task" a full set theory (p.107fn4). See section 3.3, for the potential relevance of this to contemporary minimalist syntax.

54 It would be tempting to construe this hierarchy as following a path of epistemological priority from signal to message, although some of the lay-out derives from structuralist methodology rather (cf. Chomsky 1975/1955[1], p.165).

55 Explicit reference to levels in this technical sense is made by Chomsky (1995a, p.2).

56 See the papers in Haider&Netter (eds.) (1991) for some primarily empirical discussion, and Pollard &Sag (1994) for the proposal to capture levels in terms of sortal distinctions among features.

Now, an early addition to the generative structure of language in (20) is the "lexicon." This "component" collects and organizes the vocabularies of the successive levels up to $\mathbb{P}$, i.e. phrase structure. In that sense it forms a resource for processes on practically all levels of structure, a fact that makes proper interfacing of the computational system with the lexicon difficult to formulate (cf. Chomsky 1965, 1995a, Heny 1979, Jackendoff 1997, and Bierwisch 1997).

As far as "narrow syntax" goes, only levels $\mathbb{M}$ and $\mathbb{T}$ of (20), the neighbors of phrase structure, have to concern us any further here. Already Chomsky (1975/1955[1], p.172) considered a set of "grammatically functioning morphemes," well-known elements of verbal inflection like *-ing*, *-ed*, and abstract entities like *Past* among them, to contribute to the primes of phrase structure. We're going to see how intimate a relation there continues to hold between syntax and functional morphology even in minimalist syntax.

Transformational structure, finally, is the most complex level syntacticians have to deal with. It may immediately be objected that all we want is a syntactic description of sentences, which has already been demonstrated to be achievable in terms of phrase markers or trees. Still, one of the early insights of structural linguistics concerns regularities between pairs (or groups) of sentences.

(21) a. John will leave
b. Will John leave
c. The borogoves greet the toves
d. The toves are greeted by the borogoves
e. Kermit likes beans
f. Beans, Kermit likes

"Subject-auxiliary inversion" (21a), "passivization" (21d), and "topicalization" (21f), to give just a few examples, seem to derive from an (unmarked) declarative pattern,[57] i.e. they could be formulated as functions from a declarative pattern to the respective output pattern. The exact construction of the transformational level will not be discussed here. In fact, many substantive changes introduced throughout the history of generative linguistics concern this level (see Chomsky 1965, 1971, 1981, 1995a, Gazdar et al. 1985, Heny 1979, Lasnik&Kupin 1977, Peters&Ritchie 1973).[58] We are going to struggle enough with the latest set of changes below.

---

57 More precisely, abstract underlying "kernel sentences" are "trivially" related to English declarative sentences and "non-trivially" related to transformations thereof (cf. Chomsky 1957, 1965). See Bierwisch (1963) for arguments of a similar nature concerning the V2 property of German. Given the quasi-obligatory V2 nature of German declarative main clauses, the relation between kernel sentences and declarative patterns must be taken to be somewhat more indirect in German.

58 The construction of the level $\mathbb{T}$ in Chomsky (1975/1955[1]) (cf. Chomsky&Miller 1963), though outwardly unwieldy and confusing at first sight, has a very elegant core. $\mathbb{T}$ simply allows constructs from the next lower level $\mathbb{P}$ to be among its primes. Concatenation thus operates on terminal strings (Z), phrase markers (K), and (strings of) transformational functions ($T_1 + \ldots + T_n$) to produce objects like $Z+K+T_1+ \ldots +T_n$, so-called "T-markers" (cf. Chomsky 1965, p.128ff). Of course, concatenation does not have anything to do with "temporal order of sounds" here (p.380). Instead,

Note, however, that the earliest conception of "generalized transformations" belongs here. Since, somewhat arbitrarily, no recursion was recognized at the level of phrase structure in Chomsky (1975/1955[1]), rules like (22) couldn't be put into the grammar directly. (I limit myself to recursion of S for the sake of brevity.)

(22) a. S → S+and+S
b. VP → V+S

Yet, sentences like (23) clearly belong to the object of study.

(23) a. [$_S$ [$_S$ Kermit snores] and [$_S$ Humpty Dumpty chuckles]]
b. [$_S$ Kermit [$_{VP}$ knows [$_S$ that Humpty Dumpty chuckles]]]

To produce these structures, generalized transformations were used, which can roughly be seen to be functions from pairs of phrase markers to single phrase markers. Chomsky (1965) obviated the need for this extra device by permitting phrase structure rules to be recursive, so generalized transformations were forgotten until their revival in minimalist syntax, where, as we are going to see in more detail below, they bring about a substantial approximation of the transformational and the phrase structure level.[59]

2.4 D-Structure, Checking, and the Theta Criterion

Let us next have a closer look at some essential modifications minimalism has in store for the architecture of I-language.[60] To do this, we must compare minimalist syntax with its immediate predecessor, government-binding theory (GB).[61]

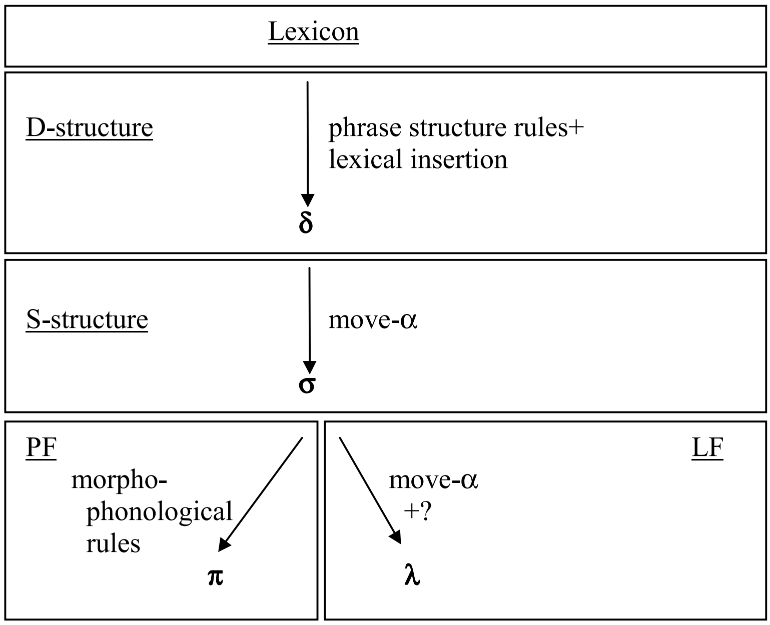
in the case at hand, it is interpreted as function application, i.e., "Z+K+T corresponds to T(Z,P)" (p.376). Clearly, however, tough restrictions have to be put on the functions allowed here, as they would overgenerate massively otherwise. Just a superficial glance makes it seem advisable to seek formalisms that remain close to the phrase structure level. Not surprisingly, this is where fierce controversy divides the field.

59 For further discussion of the early concept of generalized transformations, see Chomsky (1957), Chomsky&Miller (1963), and Fillmore (1963).

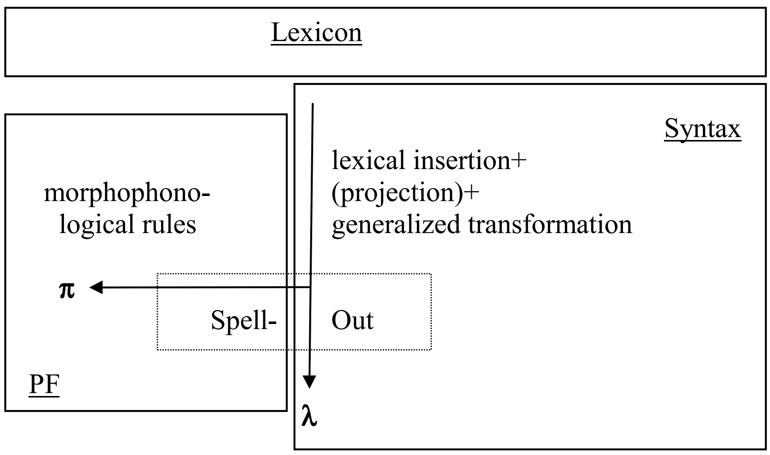
60 For the following, see Chomsky (1995a, chapter 3).

61 See Chomsky (1981, 1982, 1986a), Fanselow&Felix (1993), Grewendorf (1988), Haegeman (1992), Lasnik&Uriagereka (1988), van Riemsdijk&Williams (1986), and von Stechow&Sternefeld (1988).

(24) a. Government-Binding Theory



b. Minimalism



If anything interesting (like reconstruction for example) corresponds to the question mark at LF in (24a), a skillful theoretician could presumably bring the GB model into a format that recognizes 4(+x) levels in exactly the sense of section 2.3 above.[62]

---

[62] Clearly, D-structure could be made to correspond to level $\mathbb{P}$, S-structure to $\mathbb{T}$, and PF to the lower levels (+x) of Chomsky (1975/1955[1]).

Otherwise, S-structure and LF recognize formally identical structures and operations, which makes them indistinguishable, except for different representational conditions imposed on the respective phrase markers σ and λ.

Above, we have already seen that minimalism does away with all non-interface levels. The most reliable formal difference on which to base a distinction of levels in the syntactic domain, namely, the one between string rewriting and tree mapping, is clearly not deemed an obstacle to unifying the three levels of D-structure, S-structure, and LF. I've called the result “Syntax“ for want of a better term. The operations yielding λ in (24b) will be subject to scrutiny later.

Among the broader consequences of this minimalist design belongs the relocation of GB conditions formerly applying at δ and σ. Case theory is going to be instrumentalized for the operations putting λ and π together. Likewise, bounding theory is considered to follow from the nature of these operations. Binding (and control) theory is largely transferred to λ, though its exact effects on convergence are left open. Last but not least, theta theory, originally providing the most important condition on which syntactic elements are licensed at δ, acquired a controversial status. Theta marking imposed an order on the arguments within VP and served to define A-positions as those positions theta roles are potentially assigned at.[63] The A/A' (“non-A“) distinction played a crucial role in capturing binding and extraction phenomena as well as in the analysis of passive, raising, and ECM-constructions.

Given the usefulness of theta marking in that respect, one could expect the enforcement of theta theory to simply be transferred to λ. A theory that does this could be called “implementational minimalism.“ It would basically follow Koster (1987), who already argued for the reduction of syntactic levels to a single one. By the logic of trace theory, δ, the input to movement operations will automatically be preserved in the output representation, no matter whether the latter is taken to be σ, as argued by Koster, or λ (Chomsky 1995a, Brody 1995).[64]

However, Chomsky (1995a, chapter 3) considered a stronger alternative, which could be called “eliminativist minimalism.“ According to this alternative, violations of the theta criterion (25) wouldn't affect convergence (well-formedness as defined C_{HL}-internally) but only lead to “deviant interpretation“ (Chomsky 1995a, p.187).[65]

---

63 More precisely, phrase structure rules can be taken to define grammatical functions (GF) like “subject“ (of S), [NP,S], and “object“ of VP, [NP,VP], or “complement“ of lexical head more generally. The corresponding positions are called A-positions (A ≈ “argument“) and the associated GFs are called A-GFs. The A-positions, then, are the potential theta positions, i.e. elements in A-positions can receive theta roles (cf. Chomsky 1981, section 2.2). Technically, another principle is required to make the GB system work, namely, the “Projection Principle“ (see references cited above).

64 See Pullum (1996) for some remarks on the rationale for eliminating levels.

65 “Deviance“ is meant to indicate a weaker effect than a violation of FI, the latter being criterial of convergence (see section 2.2 above). Thus note the following passage from Chomsky (1995a, p.194): “We now say that the representation λ satisfies FI at LF if it consists entirely of legitimate objects; a derivation forming λ converges at LF if λ satisfies FI, and otherwise crashes. A convergent derivation may produce *utter gibberish*, exactly as at PF. *Linguistic expressions may be “deviant“ along all sorts of incommensurable dimensions* [. . .]“ (italics mine, H.M.G.).

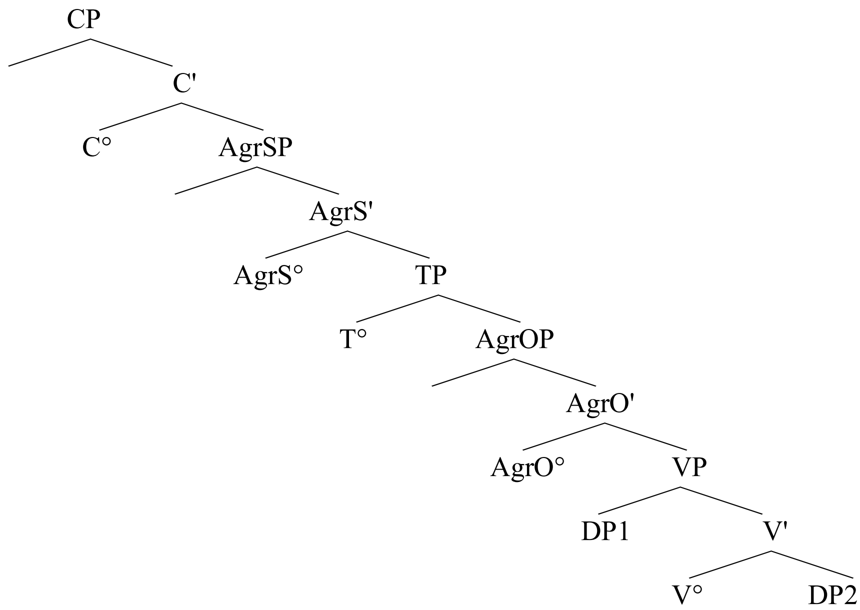
(25) Theta Criterion
Each argument bears one and only one Θ-role, and each Θ-role is assigned to one and only one argument (Chomsky 1981, p.36).

(26), for example, could then be said to be convergent but deviant (#).

(26) # John depended

Yet, deeper speculations about the quasi-semantic properties of theta theory, which tend to set in at such a point, can be skipped, since Chomsky himself provided technical arguments for keeping (at least one half of) the theta criterion as a condition on convergence. To illustrate this, I have to first introduce more of the concrete syntactic structures C_{HL} generates. The most influential version of minimalism takes (27) to be the basic universal clause structure of human language.[66]

(27)



One driving force behind postulating (27) is the fact that it allows a uniform treatment of structural Case in terms of "feature checking" in Spec-X° relations.[67] Thus, ("deep")

---

[66] Adapted from Chomsky (1995a, p.173). Linear order can vary. Sisters of X' are called "specifiers" and sisters of X° are called "complements."

[67] For relevant discussion of the merits and problems of this structure and the "Split-INFL Hypothesis" see Abraham et al. (eds.) (1996), Collins (1997), Iatridou (1990), Ouhalla (1991), Pollock (1989, 1993), and Sabel (1996), among many others. VP must probably be more richly structured (cf. Larson 1988). Auxiliaries, participles, negation and infinitival structures require additional adjustments. For the checking properties of clauses, see Bošković (1995).

subjects (=DP1) check nominative case in the specifier of AgrSP, objects (=DP2) check accusative case in the specifier of AgrOP.[68] Case checking is interlocked with the morphological properties of verbs and their functors AgrO°, T°, AgrS° (and C°), the functional heads. Universally, verbs check all the "V-features" (more neutrally "X°-features") of the functors, thereby satisfying their own morpho-syntactic requirements (agreement, tense, finiteness etc.). Language-particular differences arise, because only some of the displacements necessary for checking are "visible" at π. That's why the PF branch in (24b) isn't firmly attached. Spell-Out can apply anywhere in the bottom-up derivation from the lexicon to λ, making visible whatever string has been produced up to that point.

As already indicated in section 2.1, what decides on the fate of individual languages is the strength property of the features to be checked. Take the functional heads with two crucial features, one to be checked by its specifier ("XP-features"), one to be checked by the appropriate head, adjoining to the functional head directly ("X°-features"). For the sake of concreteness let these features be quadruples <ATT,VAL,0/1,0/1>, an "attribute" like Agr, its value, in the case of Agr itself a complex structure comprising attributes "gender, " "number, " and "person" and their respective values, plus two "bit-switches," the first one offering a choice between *0 = weak* and *1 = strong*, and the second one between *0 = checked* and *1 = unchecked*. At the functional heads some superordinate structuring must regulate which feature is available for the specifier and which for the adjoining lexical head. The items that require this morphological licensing (verbs, DPs, operators . . .) possess the appropriate featural counterparts, so that unchecked features get checked (<ATT,VAL,0/1,1> → <ATT,VAL,0/1,0>) if (i) attributes and values of both functor and argument match and (ii) the correct structural relation (spec-head, head-head) holds.

Crucially, unchecked features violate FI at λ. At π, however, only unchecked features that are *strong*, i.e. <ATT,VAL,1,1> cause a violation of FI while unchecked weak ones (<ATT,VAL,0,1>) are ignored there. Therefore, strong features must be checked *before* Spell-Out applies.[69] Determining strength individually for different functional heads in different languages, then, brings about word-order variations like the famous one in (28).[70] (*I* is the usual cover term for sentence- internal inflectional elements like *Agr* and *T*.)

[68] Ergative/absolutive languages could be taken to assign ergative Case in Spec,AgrSP and absolutive Case in Spec,AgrOP. While passive in nominative/accusative languages "absorbs" the Case in Spec,AgrOP, antipassive in ergative/absolutive languages absorbs the Case in Spec,AgrSP.

[69] This kind of mechanism has a long tradition in generative syntax. Certain dummy elements lead to ill-formedness at surface structure unless they have been operated on by appropriate transformations. See for example Chomsky (1965, p.138) and Emonds' surface filter, discussed in Leuninger (1979, p.73). My construal of FI, termed "Translatability," essentially boils down to this kind of mechanism (cf. section 3.1).

[70] See Pollock (1989). Obvious feature assignments induce the full range of typological options, that is, SVO, SOV, OSV, OVS, VSO, and VOS structures. It is not so obvious, however, whether this way of fixing the interdepencies of morphological marking and positioning of syntactic elements is on the right track. Thus, further careful analysis seems to be required for determining what amount of complementarity "marking strategies" and "positioning strategies" allow for the achievement of

(28) a. $[_{IP}$ Mary$_i$ I° $[_{VP}$ often $[_{VP}$ t$_i$ kisses John]]]
b. * $[_{IP}$ Mary$_i$ $[_{I°}$ kisses$_j$] $[_{VP}$ often $[_{VP}$ t$_i$ t$_j$ John]]]
c. * $[_{IP}$ Marie$_i$ I° $[_{VP}$ souvent $[_{VP}$ t$_i$ embrasse Jean]]]
d. $[_{IP}$ Marie$_i$ $[_{I°}$ embrasse$_j$] $[_{VP}$ souvent $[_{VP}$ t$_i$ t$_j$ Jean]]]

Assuming that French I° has a strong feature for the verb to check, while English I° is weak, the strings resulting from (28a)/(28d) are the ones that get pronounced at π. At λ, structures are assumed to be identical for all languages.[71] Given that weak features too must be checked, V°-to-I° will occur in English after Spell-Out. So, in fact, the asterisks in (28) represent the appropriate ban on output at π only.

Although we know now why *overt* V°-to-I° is forced in French, namely, because an unchecked strong feature isn't tolerated at π, there is no obvious way to ban the same operation in English from applying before Spell-Out. FI would be satisfied either way. This question is

> "answered by a natural economy condition: LF movement is "cheaper" than overt movement (call the principle *Procrastinate*). [. . .] The intuitive idea is that LF operations are a kind of "wired in" reflex, operating mechanically beyond any directly observable effects. They are less costly than overt operations. The system tries to reach PF "as fast as possible," minimizing overt syntax (Chomsky 1995a, p.198)."[72]

This type of "least effort" principle introduces a "global" or "transderivational" dimension to C_{HL}, as will be clear later on (cf. section 2.8). An *ad hoc* formulation to capture the case at hand would be (29) (cf. section 2.2 above).

(29) A derivation D_i is *well-formed* iff D_i yields an expression $<\pi,\lambda>$ that satisfies FI, and there is no derivation D_j ($j \neq i$) yielding $<\pi,\lambda>$ such that D_j requires fewer operations before Spell-Out than D_i.[73]

---

which purposes (e.g. identification of arguments or illocutionary potential, and orientation in discourse). See for example the challenging sets of facts discussed in Baker (1996), Kiss (ed.) (1995), and Speas (1990).

[71] Chomsky (1995a) locates linear ordering at PF. Otherwise, identity is restricted to hierarchical structure, i.e. the dominance relation (cf. section 2.3).

[72] Note incidentally the allusion to "mental computations" hidden in this formulation, playing on hardware/software analogies and the associated costs of design and operation. See section 1.1 above and references cited there.

[73] See Sternefeld (1997) for detailed discussion of the kind of metrics and "reference sets" involved here. I have already indicated in section 2.2 that I will advocate a version of syntax (to be introduced in section 3) that doesn't have recourse to economy principles. Such principles clearly do not belong to the elementary steps in theory construction in the sense of section 1.1 above and must therefore be independently motivated. Also, a major caveat has to be added wrt the foregoing discussion. No fully worked out checking theory has been provided by Chomsky (1995a). Rather, a number of options leading to subtle empirical and theoretical differences are discussed. I've only sketched a version that captures the basics. For a much more comprehensive discussion see for

We are now ready to again pick up the question what status the theta criterion (25) should be accorded. Checking theory as developed so far regulates the operations in the functional domain. Nothing, however, is said wrt the structure of VP, or other lexical projections like NP and AP. More precisely, the question arises why the verb and its arguments couldn't be randomly inserted somewhere in the functional domain.[74] Insertion into strong positions wouldn't be detectable at π, given that elements tend to end up there anyway. So (30), in which the subject has been directly inserted in the specifier of AgrSP, is not an immediate problem.[75]

(30) $[_{CP}$ att $[_{AgrSP}$ Maria AgrS° $[_{TP}$ T° $[_{AgrOP}$ AgrO° $[_{VP}$ hälsar Anna]]]]]
that Mary greets Ann

Weak positions, however, make a difference. Thus (31), dispensing with VP entirely, diverges from π-observable word order in Swedish, a rather unwelcome result for syntactic theory.

(31) * $[_{CP}$ att $[_{AgrSP}$ Maria AgrS° $[_{TP}$ T° $[_{AgrOP}$ Anna $[_{AgrO°}$ hälsar]]]]]

Appeal to number of operations doesn't help, since at least one insertion has to occur in any case. Disallowing checking for insertion operations, however, would require revision of the standard analysis of expletives, which, it seems, get inserted in the functional domain chiefly for checking purposes. Thus, if (25) were a convergence criterion, lexical projections could be taken to be the domain of theta assignment, which would force arguments to be inserted in a projection of their theta licenser. Such a

example Wilder&Ćavar (1994), who conjectured that only functional heads could carry strong features, all others being weak (cf. Nunes 1995, p.312). I will come back to details of checking only where this contributes to the intelligibility of the discussion, assuming otherwise that open problems will be sorted out *somehow*. Still, I would like to add two more general points.

(i) A full-fledged theory of features is likely to have to invoke *feature structures*, which incidentally a look at the *feature tree* and the issue of *feature geometry* in phonology would seem to confirm (Kenstowicz 1994, section 4.3). In the area of feature structures, a lot of work has been done by e.g. Gazdar et al. (1985), Johnson (1988), Pollard&Sag (1987, 1994), and Shieber (1986), developing a representational format (and formalisms) that enhances the transparence of the kind of discussion minimalism is getting involved in because of its heavy reliance on features and lexical specifications. See Wartena (1994) for a unified perspective.

(ii) Allowing feature checking to result in the "turn of a switch" (<ATT,VAL,0/1,1> → <ATT,VAL,0/1,0>), offers "representational" options, if necessary. See Stabler (1996) for similar mechanisms. The Chomskyan system assumes deletion to follow checking operations. This raises further difficult issues, leading among other things to a further split between "deletion" and "erasure" (Chomsky 1995a, p.280ff). Both theories share the property of providing only finite resources for syntactic operations, if the latter are essentially reduced to checking or cancelation operations (cf. Stabler 1996, 1998) and Michaelis (2001).

[74] Cf. Chomsky (1995a, p.314ff).

[75] Data from Swedish illustrate my point most directly.

conclusion is also reached by Chomsky (1995a, p.316), if I understand that correctly.[76] I will thus formulate the ("conditional") hypothesis 2:

Hypothesis 2: Under certain conditions, (part of) the theta criterion has to be implemented as a syntactic well-formedness constraint (=H2)

The conditions that call for such an implementation are discussed in sections 3.4.2 and 3.5 below.[77]

2.4.1 Idioms (An Excursus)

Eliminating D-structure would seem to have another consequence not yet fully appreciated in expositions of minimalist syntax. The analysis of idioms, earlier assumed to essentially have recourse to D-structure,[78] has to be reconsidered. Crucially, a lot of idioms cannot undergo transformations like passivization.

(32) a. John kicked the bucket
b. The bucket was kicked by John

Only (32a) can mean that John died while (32b) is restricted to its literal meaning. Idioms that allow transformations constitute a subset of the class of idioms, by and large.

> "Thus idioms in general have the formal properties of non-idiomatic structures, and appear either in D-structure or S-structure form, but not only in S-structure or LF-form. D-structure not S-structure or LF, appears to be the natural place for the operation of idiom rules, since it is only at D-structure that idioms are

[76] The second part of (25) seems to be implied by the following formulation: "[. . .] a transitive verb assigns a Θ-role *by definition*" (ibid., italics mine, H.M.G.). The first half of (25), though, is only fully adopted as a convergence criterion, where expletive and ECM constructions are analyzed (Chomsky 1995a, section 4.9).

[77] See Higginbotham (1985) and Speas (1990) (cf. Bierwisch 1997), for an explicit option, which additionally provides mechanisms for an adequate treatment of adjunction. A strong defense of theta theory will, of course, have to show that it can be raised above the level of "pseudosemantics" (Dowty, 1989). In fact, different kinds of saturation mechanism could be understood as an implementation in the spirit of H2. Thus, subcategorization or c-selection seem to be sufficient and appropriate for the ordering of functional categories. See also Dowty (1982) for a proposal based on alternative views concerning the interaction of syntax and semantics. For the purposes discussed in sections 3.4.2 and 3.5, an arbitrary association of lexical items with selectional grids would formally be sufficient.

[78] "These properties of idioms provide evidence in support of an independent syntactic component mediating the relation between sound and meaning in terms of D- and S-structures, and also in support of the *existence of D-structures* as distinct from S-structures, related to the latter by grammatical transformation" (Chomsky 1980a, p.153; italics mine, H.M.G.).

> uniformly not "scattered" and it is only the D-structure forms that always exist for the idiom (with marked exceptions), S-structure sometimes being inaccessible to idiomatic interpretation. Thus at D-structure, idioms can be distinguished as subject or not subject to Move-α, determining the asymmetry just noted" (Chomsky 1981, p.146fn.94).

What the exact formulation of idiom rules would be isn't quite clear.[79] As Jackendoff (1997) points out, all idioms would have to be subject to a considerable number of "minor" movement rules in minimalist analyses in any case.

Now, except for reconstruction effects, idioms are only briefly mentioned in Chomsky (1995a, p.207).

> "Having abandoned D-Structure, we must assume that idiom interpretation takes place at LF, as is natural in any event."

Although intuitions as to what appears natural sometimes vary on theory-internal grounds, there are additional facts that do argue for an even fuller dissociation of syntactic and semantic aspects of idioms. Thus, consider (33).

(33) $[_{DP}$ The headway $[_{CP}$ John made]] was surprising

Given that an idiomatic construal of *make headway* is possible here, examples like this have been taken to require a "head-raising" analysis of relative clauses, such as recently defended by Kayne (1994). Accordingly, *headway* would have undergone movement to its surface position, compatible with a "contiguous" D-structure representation of the idiom parts inside the relative clause, as sketched in (34).

(34) $[_{DP}$ The $[_{CP}$ John made headway]]

The stronger claim that head-raising is the canonical analysis of relative clauses, however, gets into conflict with exactly the kind of data it was supposed to be accounting for.[80]

---

[79] Noteworthy in the context of this dissertation is the fact that Chomsky (1981, p.146fn94) played with the idea of a solution in terms of the string-set theory of phrase markers (Chomsky 1975/1955[1], Lasnik&Kupin 1977). Thus, (32a) would be represented by phrase marker (i) (terminals omitted).
(i) {S, NP+VP, NP+V+NP, NP+V+Det+N, **NP+V**}
Adding *NP+V*, a technique later become known as "reanalysis" (cf. Haegeman&van Riemsdijk 1986), represents the fact that *kick the bucket* must be considered a verb at the appropriate level of analysis.

[80] For the following data as well as further insightful discussion, see McCawley (1981), McCloskey (1979), and Nunberg et al. (1994).

(35) a. John made friends with Bill
b. Friends can be made with anybody
c. * The friends that John made with Bill was/were lasting
d. Parky pulled the strings that got me my job

Although *friends* can be detached from its idiomatic partner *make* without affecting interpretation in (35b), (35c) isn't licensed. Even worse, (35d) requires *pull strings* to be base-generated in “scattered“ position under the head-raising analysis of relative clauses. Nevertheless, idiomatic interpretation of (35d) is straightforward. Not even the proposal to apply head-raising only optionally would suffice to defend a D-structure account of idioms, as (36) indicates (McCloskey 1979, attributed to Joan Bresnan).[81]

(36) We made [what later seems to have been described as great headway] on this problem

Thus, idiom interpretation must be able to proceed in a more abstract fashion. Syntactic contiguity of formatives is not required.[82] In that respect, the minimalist elimination of D-structure is fully vindicated.

[81] Chris Wilder (p.c.) points out that (36) might possess a bracketing that locates *headway* outside of the free relative clause. This, however, neither corresponds to my semantic intuitions nor can it be upheld when (36) gets translated into German.

(i) . . . weil wir [was später als großer Fortschritt beschrieben wurde] machten

Due to the SOV nature of German *Fortschritt*, the counterpart of *headway*, is clearly determinable as being placed inside the free relative.

[82] For relative clauses, optional head-raising could be dispensed with in favor of “garden-variety“ operator or pronoun analyses. The following data suggest that pronouns must be able to transfer or sustain idiomatic interpretation anyway.

(i) Once someone lets the cat out of the bag, it's out of the bag for good (Nunberg et.al. 1994)

(ii) Den Vogel den hat Ralf abgeschossen (Ralf Vogel, p.c.)
The bird that has R. shot down
“Ralf took the cake“

Thus, *it* in (i) refers to the proverbial cat. Furthermore, (ii) is standardly analyzed as left-dislocation, which means that only the demonstrative *den* can be inserted as complement to the idiomatic partner of *Vogel*, i.e. *abgeschossen*. See Jackendoff (1997) for a similar assessment on different grounds and for far-reaching steps toward an alternative theory.

2.5 Minimalist Operations I

Sections 2.5 and 2.6 will be devoted to having a closer look at the syntactic operations employed in minimalist syntax. Major change is already adumbrated in (24) above, where the overall architecture of minimalism is laid out. Recall that minimalist syntax was initially developed in two stages, as is now recorded in chapters 3 and 4 of Chomsky (1995a). I will retrace that order of events, gradually introducing the core issues of this study.[83]

2.5.1 Project-α, Move-α, and Generalized Transformation

With the minimalist elimination of D-structure and its designated level marker δ, phrase structure rules have been left "dangling in the air," so to speak. This at least follows if generating δ was their sole purpose. We have already seen in section 2.3 that the picture is more complicated than that. Thus, even λ (and the eliminated σ) is a phrase marker which displays properties definable in terms of phrase structure rules.[84] Thus, abandonment of D-structure only means that the function of these rules has to be reassigned. How this is achieved will be discussed shortly. Chomsky (1995a) gives additional empirical evidence that something is wrong with the GB account of D-structure in any case. Consider (37).

(37) a. It is easy to please John.
b. John is easy [$_{CP}$ OP$_i$ [$_{IP}$ PRO to please t$_i$]].
c. [That economy was difficult to control] is easy to see.

The expletive *it* in (37a) is a criterion for assuming that the matrix subject position is not a theta position. Secondly, operator movement from the embedded object position in (37b), taken as the proper analysis of such constructions on independent grounds (cf. Chomsky 1977, 1981), forces insertion of *John* in a non-theta position (Chomsky 1995a, p.188). This kind of insertion must occur after δ has been assembled, given that δ is seen as a pure representation of theta marked grammatical functions. Further, (37c) shows that post D-structural insertion wouldn't just be a matter of accessing the lexicon. Instead, constructing the unboundedly complex items that can later be inserted requires the full power of phrase structure rules and transformations as well. To save the GB picture one could have δ to be a set of (sub-)phrase markers such as (38).

83 Meanwhile the program has undergone further revision, as documented in Chomsky (2000, 2001). I will make reference to such revisions only where they fundamentally concern the results arrived at here. See especially section 2.7.4 below.

84 See for example Gazdar et al. (1985).

(38) δ = { $[_{NP}$ economy],
$[_{CP1}$ That $[_{IP}$ e $[_{VP}$ was difficult $[_{CP}$ $[_{IP}$ PRO to control Op]]]]],
$[_{CP2}$ $[_{IP}$ e $[_{VP}$ is easy $[_{CP}$ $[_{IP}$ PRO to see Op]]]]] }

At S-structure, two kinds of process are then required, namely, (i) operator movement to the specifiers of CP in both CP1 and CP2, as well as (ii) insertion of the NP in CP1 and of the completed CP1 into CP2. The type of operation in (ii) is called a "generalized transformation" in the sense of Chomsky (1975/1955[1]) (see section 2.3 above). In the face of these facts, it is legitimate to conclude that, with the abandonment of D-structure, objects like δ in (38) get a more natural interpretation. They constitute an intermediate stage in a derivation from the lexicon to λ, without possessing any designated representational status.[85]

Consequently, Chomsky (1995a) proposes to base all of syntactic derivation on the free application of three operations, "Project-α" (39), "Singulary Transformation" or "Move-α" (40), and "Binary-" or "Generalized Transformation" (GT) (41) (Chomsky 1995a, p.189).[86]

(39) Project-α
a. selects an item α from the lexicon and projects it to either i., ii., or iii.
i. $[_{X^\circ}\ \alpha\ ]$
ii. $[_{X'}\ [_{X^\circ}\ \alpha\ ]\ ]$
iii. $[_{XP}\ [_{X'}\ [_{X^\circ}\ \alpha\ ]\ ]\ ]$

---

85 For discussion of (37) see also Frank&Kroch (1995). The exact analysis of (37) is not as clear as it would seem from (Chomsky 1995a). Thus, Chomsky (1986a, p.114) states that the matrix subject position of (37b) in fact *is* a theta position. In the light of the analysis in Chomsky (1981, section 5.4), I take that to mean that the matrix subject position acquires a [+ theta] status only after operator movement has taken place within the infinitival CP, possibly followed by reanalysis of the adjective and its complement (Chomsky 1981, p.312). Anyway, an S-structural operation, move-α, would have to precede a putative D-structural one, lexical insertion. Giving up these levels resolves the looming paradox. See Wilder (1991) for further discussion and an alternative analysis.

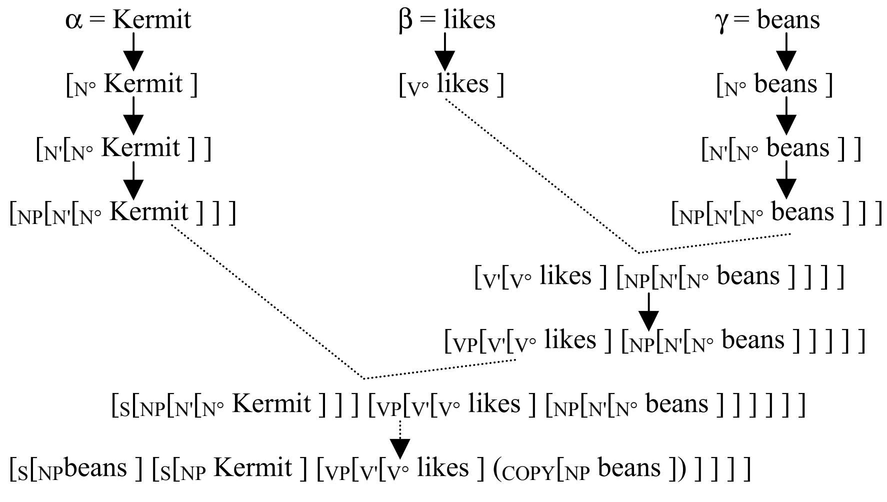
86 This system closely resembles the proposal in Speas (1990). It seems, though, that in her system, concatenation, as expressed by GT, is left undefined (cf. Zwarts 1992, p.20). See below for discussion of Lebeaux (1991). The programatic nature of Chomsky (1995a, chapter 3) led to a rather complex intersection of perspectives. Several of the concepts appealed to possess definitions only in terms rendered obsolete by the discussion itself. This quite naturally provoked the more radical theory of Chomsky (1995a, chapter 4), which will be investigated in section 2.6. I therefore refrain from a more comprehensive analysis of this intermediate stage here (cf. Wartena 1994), just trying to present a "view from nowhere."

(40) Singulary Transformation (= Move-α)[87]
ST(K) = K*, for K and K* phrase markers that satisfy X-bar theory
a. subroutines of ST
  i. Target K
  ii. Add ∅
  iii. Substitute α for ∅, where α is a phrase marker within K. Leave a copy of α in the original position of α. Form the chain (α, ($_{COPY}$ α)).
  iv. Form K*
b. sample application
  i. $[_K\ \alpha\ ] \rightarrow$ ii. $\varnothing\ \ [_K\ \alpha\ ] \rightarrow$ iii. $\alpha_i\ \ [_K\ (_{COPY}\ \alpha_i)\ ] \rightarrow$ iv. $[_{K^*}\ \alpha_i\ [_K\ (_{COPY}\ \alpha_i)\ ]\ ]$

(41) Binary Transformation (= Generalized Transformation)
GT(K,K') = K*, for K, K', and K* phrase markers that satisfy X-bar theory
a. subroutines of GT
  i. Target K
  ii. Add ∅
  iii. Substitute K' for ∅
  iv. Form K*
b. sample application
  i. $K \rightarrow$ ii. $K\ \ \varnothing \rightarrow$ iii. $K\ \ K' \rightarrow$ iv. $[_{K^*}\ K\ K'\ ]$

(42) gives a simplified derivation on the basis of these operations (Project-α = ⟶, Move-α = ⋯▶, GT = ⋙).[88]

(42)



[87] Chain formation is indicated by coindexation in (40b). See sections 2.7 and 3.3 for the complexity of chains and the apparent necessity of using something like indexation to capture their function.

[88] The careful reader will realize that I've adapted Project-α in so far as I've applied it recursively, which means that phrase markers have to be in its domain as well. Also, NP-internal projections are left out in the final step to enhance readability.

The resemblance between (42) and transformation markers (Chomsky 1965, p.130) is intended, given that the revival of generalized transformations taps resources from "standard theory." To allow free interspersing of rules, only insufficiently illustrated in (42) and fully emerging in later sections, incidentally means to present a grammar the properties of which were implied by an early conjecture of H.B. Curry's (1961, p.65f).

> "From the standpoint of tectogrammatics I see no reason to put phrase structure and transformation grammar on separate levels, nor to suppose that phrase structure operations necessarily either precede or follow transformation operations."

"Tectogrammatics" is an abstraction from "phenogrammatics," the latter responsible for capturing "surface" aspects of expressions like linear precedence.[89]

Now, it is clear from the explicit restrictions on GT and Move-α as well as from the notations used in Project-α that everything is still held together by *X-bar theory*. The latter is usually given in the form of phrase structure rule schemata. (Variables W, X, Y, Z range over category labels, α ranges over lexical formatives.)[90]

(43) X-bar Theory

a. $XP \rightarrow XP + Y^?$	b. $XP \rightarrow Y^? + XP$	c. $XP \rightarrow (Z^?) + X'$
d. $XP \rightarrow X' + (Z^?)$	e. $X' \rightarrow X° + (W^?)$	f. $X' \rightarrow (W^?) + X°$
g. $X° \rightarrow W° + X°$	h. $X° \rightarrow X° + W°$	i. $X° \rightarrow \alpha$

(43) guarantees "endocentricity" of syntactic phrases (XPs), that is, there is an unbroken sequence of projections <XP, X', X°> from each phrase to its "head" (X°).[91] We can say, then, that phrases and their heads have a common distribution, differing only in the degree of "saturation," as this would be called if heads were seen as functors. Optionality of complements and specifiers, indicated by parentheses, implies that trivial functors are permitted. Adjuncts of XP ($Y^?$ in (43a)/(43b)), specifiers ($Z^?$ in (43c)/(43d)), and complements ($W^?$ in (43e)/(43f)) are usually taken to be maximal phrases as well. I'll come to this shortly.

Given the bottom-up character of syntax suggested by abandoning D-structure and invoking GT, X-bar schemata look like a superfluous addition to the grammar. Indeed, if (43) were made into a collection of tree formation operations,[92] these could do the job of Project-α as well as being "called as module" by GT and Move-α. X-bar rules would have to be interpreted from right to left then. Secondly, decisions would have to be made on the domain of projection and concatenation,[93] i.e. whether they operate on

---

[89] See also Dowty (1982).

[90] For a comprehensive discussion, see Kornai&Pullum (1990).

[91] Thus, rules like S → NP+VP, which I'm occasionally using as a shortcut, aren't allowed and must be replaced (cf. section 2.4).

[92] Cf. Partee et al. (1993, p.444) and Manaster-Ramer&Kac (1990, p.351).

[93] In their classical form, generalized transformations would, again I'm simplifying, be defined over an ordered pair of ordered pairs $\langle\langle Z_1,K_1\rangle, \langle Z_2,K_2\rangle\rangle$ and produce a new ordered pair $\langle Z_3,K_3\rangle$,

lexical formatives, nodes, categories, trees, or other objects still. However, I won't go into details here, given that a more radical reform of syntax lies still ahead.

Note also that the reference to phrase markers in the formulation of GT and Move-α is nontrivial, if the term "phrase marker" were taken in its strictest sense as a level-marker of level ℙ, or its successor, the now obsolete level of D-structure. In a looser sense, there still remains the choice between stringsets and trees.[94] Nevertheless, I would like to pinpoint the following minimal properties of Project-α, GT, and Move-α, expressible in terms of trees.

Project-α defines transitions from lexical formatives, which initially equal zero nodes for technical reasons, to objects made up of two, three, or four nodes. GT is a function from pairs of objects possessing *n* and *m* nodes respectively to objects of $n+m+1$ nodes. Move-α, finally, is a function from objects of *n* nodes to an object of $n+m+1$ nodes, where *m* is the number of nodes possessed by the object which moves ($m \leq n$). Call this the "node record" of these operations.[95]

(44) Node Record

a. Project-α (0)	→	m	($2 \leq m \leq 4$)
b. GT (n,m)	→	n+m+1	($n > 0, m > 0$)
c. Move-α (n)	→	n+m+1	($n > m > 0$)

Again, the similarity between binary and singulary transformations should not be overlooked. They differ, though, in the number of "source objects," given that only maximal phrase markers are counted as input to these operations. The reference to sub-

where Z is a terminal string, K is a phrase marker, i.e. a set of strings such that $Z_i \in K_i$. Z_3 then is the result of concatenating Z_1 and Z_2, (Z_1+Z_2), and K_3 "is essentially the Cartesian product of K_1 and K_2 (i.e. it is the set of strings that contains X_1+X_2 wherever $X_1 \in K_1$ and $X_2 \in K_2$)" (Chomsky 1975/1955[1], p.383). An adapted example would look like (i) and (ii).

(i) Z_1 = the+borogoves, Z_2 = greet+the+toves, Z_3 = the+borogoves+greet+the+toves

(ii) K_1 = {NP, Det+N}, K_2 = {VP, V+NP, V+Det+N}, K_3 = {S, NP+VP, NP+V+NP, NP+V+Det+N, Det+N+VP, Det+N+V+NP, Det+N+V+Det+N }

I've smuggled *S* into K_3 and left out the terminals, which belong inside phrase markers as well, for the sake of brevity. As already mentioned in section 2.3, in this earlier model, generalized transformations didn't replace but complement phrase structure rules, thereby mainly providing recursion on S. A contemporary theory of grammar using this kind of generalized transformation in factoring recursion (and dependencies), is "tree adjoining grammar" (TAG) (Abeillé 1993, Frank 1992, Joshi 1985).

94 The following remark doesn't make a strong commitment to trees very likely. "Each level is a symbolic system, consisting of atomic elements (primes) and objects constructed from them by concatenation and other operations. We take these objects to be phrase markers in the familiar sense (*represented conventionally by trees or labeled bracketing*)" (Chomsky [with Howard Lasnik] 1995a, p.34; italics mine, H.M.G.).

95 Adding *m* in (44c) is necessary because Move-α creates a full copy of the object displaced. The node record can be useful in keeping track of items in adjunction structures (cf. section 2.6). Likewise, a closely related notion can function in proofs about graphs (cf. section 3.3).

phrase markers in Move-α will later (cf. section 2.6) be replaced by a more principled distinction, so I will postpone closer attention to source objects till then.

Finally, well-known open questions of X-bar theory carry over from GB. An attempt to achieve a uniform picture is presented by Wartena (1994, p.18), who proposed to have Project-α create $[_{XP} [_{X'} [_{X°}\ \alpha\ ]\ ]\ ]$ obligatorily. This would obviate the need for GT and Move-α to fill in extra nodes. Unfortunately, adjunction stands in the way of this simplification, at least to the extent that it's application can't be foreseen on the basis of feature structures of projecting lexical items. Otherwise, an element of optionality would have to be reintroduced into the formulation of Project-α.

In fact, Chomsky's exposition implies a rather more liberal view on X-bar theory. Where earlier versions of the theory clearly required both specifiers and complements to be maximal phrases (=XP) (e.g. Chomsky 1986b, p.3), the embedded I' node in (45) is allowed to violate that condition (Chomsky 1995a, p.190).

(45) $[_{I'}$ seems $[_{I'}$ is certain [John to be here]]]

However, the more detailed discussion of transformations contains the following passage (Chomsky 1995a, p.190).

> "For example, we can target K = V', add ∅ to form $[_{\beta}$ ∅ V'], and then either raise α from within V' to replace ∅ or insert another phrase marker K^1 for ∅. In either case the result must satisfy X-bar theory, which means that the element replacing ∅ must be a maximal projection YP, the specifier of the new phrase marker VP = β."

This seems to allow two interpretations. Elements replacing ∅ must always be maximal or specifiers must always be maximal. To be consistent with (45), the latter interpretation has to be chosen. As a consequence, structures like (46) would be allowed in complement position.

(46) $[_{D'}$ the $[_{N'}$ shooting $[_{P'}$ of $[_{D'}$ the $[_{N°}$ hunters]]]]]

On raising to a specifier, e.g. for Case checking, Project-α would have to adjust the X-bar status of D' to DP. I'm indulging in this piece of pedantry, because the unanswered questions of X-bar theory seem to rearise even in minimalist approaches.

Q2: (i) Are (dummy) specifiers obligatory?
(ii) Is non-branching projection obligatory for X' in the absence of a specifier?
(iii) Can Project-α adjust elements when they become specifiers?
(iv) Can intermediate projections be skipped?[96]
(v) Can X-bar theory be derived?[97]

---

[96] Cf. Chomsky (1986b, p.4).

[97] Various attempts in this direction have been made, the most important recent one by Kayne (1994).

We are going to see in section 2.6 that Chomsky (1995a, chapter 4) attempts to do away with the most part of these rather artificial questions for good. Q2(v), i.e. the theoretically most interesting question, will, however, be left open.

2.5.2 Constraints on Transformations: Cyclicity

Let's turn to some more empirical issues. The unordered application of transformations raises questions of locality, which occupied pre-minimalist generative syntax to a considerable degree.[98] Consider derivations of the following type, discussed by Chomsky (1995a, p.190).[99]

(47) a. $[_{C'}$ did $[_{IP}$ John wonder $[_{C'}$ C° $[_{IP}$ Mary fixed what how]]]]
b. $[_{CP}$ How$_i$ $[_{C'}$ did $[_{IP}$ John wonder $[_{C'}$ C° $[_{IP}$ Mary fixed what t_i]]]]]
c. $[_{CP}$ How$_i$ $[_{C'}$ did $[_{IP}$ John wonder $[_{CP}$ what$_j$ $[_{C'}$ C° $[_{IP}$ Mary fixed t_j t_i]]]]]]

The possibility of filling the specifier of the matrix CP before the embedded one allows an adjunct to be extracted in (47b) from what later becomes a WH-island in (47c). Seeking a derivational solution to this problem would seem to imply that the two operations have to be ordered the other way round. For this purpose, Chomsky postulates a "cyclicity" requirement on operations, called "extension condition" (Chomsky 1995a, p.190).[100]

(48) Extension Condition on Transformations
For overt substitution operations, ∅ must be external to the targeted phrase marker K

The formulation presupposes that, as already mentioned above, source objects are maximal phrase markers.

Note that there are two kinds of exception to (48). First of all, covert operations like "object-shift" in English (cf. section 2.4) don't apply at the root node of their targeted

---

98 See for example Chomsky (1986b), Lasnik&Saito (1992), Hornstein&Weinberg (1995), Müller (1995), Rizzi (1990), and Sternefeld (1991).

99 I continue to use traces instead of copies where nothing hinges on the difference.

100 This condition is a strengthening of the "Strict Cycle Condition" formulated by (Freidin 1978, p.520).
(i) Strict Cycle Condition
No rule can apply to a domain dominated by a cyclic node A in such a way as to affect solely a proper subdomain of A dominated by a node B which is also a cyclic node.
Usually, S and NP were considered to be cyclic nodes. The Extension Condition on Transformations could, for the sake of comparison, be approximated by a "Stricter Cycle Condition."
(ii) Stricter Cycle Condition
No overt substitution operation can apply to a domain dominated by a node A in such a way as to affect solely a proper subdomain of A dominated by a node B.

phrase marker. Instead, ∅ is added into a position already dominated by other nodes, as illustrated in (49).

(49) a. [$_{AgrSP}$ Kermit$_i$ AgrS° [$_{AgrOP}$ ∅ AgrO° [$_{VP}$ t_i likes beans]]]
b. [$_{AgrSP}$ Kermit$_i$ AgrS° [$_{AgrOP}$ beans$_j$ AgrO° [$_{VP}$ t_i likes t_j]]]

Secondly, head movement by adjunction, whether overt or covert, has to violate (48) in principle. Thus consider (50).

(50) a. [$_{X'}$ ∅ [$_{X°}$ α] [$_{YP}$ [$_{Y°}$ β]]]
b. [$_{X'}$ [$_{X°}$ [$_{Y°}$ β]$_i$ [$_{X°}$ α]] [$_{YP}$ t_i]]

Given that [$_{X°}$ α] doesn't include any other categories, it must, for anything to be able to adjoin to it by movement, be embedded in a larger structure. Thus, ∅ has to be inserted internal to [$_{X'}$ . . .], although the latter structure counts as targeted phrase marker K for the singulary transformation that brings about head movement. On the assumption, however, that X°-movement invariably results in *adjunction*, (48) isn't violated, given that it is restricted to “substitution“ operations, the somewhat misleading cover term for introducing either complements or specifiers.[101]

Now, this second type of exception to (48) is supposed to have more than theory-internal motivation. Based on examples like (51), Chomsky (1995a, p.204) argues that a system incorporating GT in the way it is defined above can handle such cases in an illuminating way.

(51) a.* [[Which claim [that John$_i$ was asleep]] was he$_i$ willing to discuss t]
t = ($_{COPY}$ [Which claim [that John$_i$ was asleep]])
b. [[[Which claim] [that John$_i$ made]] was he$_i$ willing to discuss t]
t = ($_{COPY}$ [Which claim])

The account, earlier presented in Lebeaux (1991), is based on the assumption that adjuncts can be freely inserted into tree structures by GT at any point of the derivation. Thus relative clauses like the one in (51b), treated as adjuncts traditionally, *need not*, while complements as the one in (51a) *must* occur in base-positions. Therefore the former *need not*, while the latter *must* be part of the copy left behind by WH-movement. At LF, (51a) violates binding condition C while (51b) doesn't, on the assumption that WH-copies are visible for such a condition.

Although this is an interesting account, I think it should not be maintained, not the least because it seems to run into empirical problems.[102] Thus, it looks as if the complement clause in (52a), which has to be reconstructed, doesn't induce a condition C

---

[101] Watanabe (1995) presents a theory of cyclicity, which, in order to allow these two exceptional cases, has recourse to a global principle, banning countercyclicity only “where possible.“

[102] In fact, the follow-up system of Chomsky (1995a, chapter 4) doesn't allow counter-cyclic GT at all. The reconstruction data at issue here are clearly not seen as an obstacle to that revision. See section 2.6 below.

violation in spite of the coindexed pronoun and proper name. Still, as shown by (52b), something like condition C is clearly operative in German.

(52) a. [Wessen Behauptung [dass Maria$_i$ krank sei]]$_j$ hat sie$_i$ (durch ihre$_i$
Whose claim that Mary ill be has she through her
Anwesenheit) t$_j$ widerlegt
presence refuted
"Whose claim that Mary was ill did she refute (through her presence)"
b. * Sie$_i$ widerlegte Peters Behauptung [dass Maria$_i$ krank sei]

Postulating an adjunct/complement dichotomy appears to be insufficient to cover the data. Furthermore, Heycock (1995) presented another set of facts that go counter to the above treatment of adjunction as well.

(53) a. [Which allegations about John$_i$]$_j$ do you think he$_i$ will deny t$_j$
b. * [How many stories about Diana$_i$]$_j$ is she$_i$ likely to invent t$_j$
c. * [How many lies [aimed at exonerating Clifford$_i$]]$_j$ is he$_i$ planning to come up with t$_j$

The contrast exhibited by (53a)/(53b) causes a dilemma for reconstruction whatever the analysis of the *about-PP* turns out to be. Both sentences should behave alike. In addition, (53c) surprisingly forces an adjunct to reconstruct.

The latter observation allows me to look at the logic of the above proposal more closely. There are four cases to consider wrt to the ordering of adjunction operations, illustrated in (54).

(54) Ordering of adjunction and its effect on grammaticality

	cyclic	counter-cyclic	prediction
a.	√	√	√
b.	√	*	√
c.	*	√	√
d.	*	*	*

According to Lebeaux (1991, p.218), ungrammaticality is only predicted (prediction = *) where neither cyclic not counter-cyclic application of adjunction could rule in (√) the analysis. To show that counter-cyclic operations are welcome, one has to produce a fact that conforms to the pattern in (54c). This has been done by offering (51b). The ungrammaticality of (53c) results in a direct counterexample, given that counter-cyclic adjunction should have prevented violation of binding condition C. That prediction is not borne out. Of course, principle C effects have sometimes been considered part of a

theory of interpretation.[103] So the proper formulation of the extension condition on transformations (48) should probably not have to stand or fall with that issue.[104]
Indeed, there are cases of a different kind that, depending on their analysis, also argue against the option of counter-cyclic application of adjunction operations. The following contrast, to begin with, has been pointed out by Culicover (1993).

(55) a. * Who$_i$ do you think [that t_i will vote for Smith]
b. Who$_i$ do you think [that [after all things are considered] t_i will vote for Smith]

The well-known *that*-trace effect disappears as soon as an adjunct immediately follows the complementizer. Under a derivational account of ECP violations that marks traces as ungrammatical in the moment of chain formation, the adjunct would have to be inserted cyclically (or at least before WH-movement occurs). (55b) can be subsumed under pattern (54b), which means there is at least good reason to allow the "trivial" ordering of operations. Turning to (56), we are confronted with additional potential problems for prediction (54c).

(56) a. * I asked [what$_i$ [to Leslie]$_j$ Robin gave t_i t_j]
b. Worüber$_i$ hat keiner [ein Buch t_i] gelesen
What-about has no-one a book read
"About what did no one read a book"
c. * Worüber$_i$ hat [ein Buch t_i]$_j$ keiner t_j gelesen

Thus, if topicalization is taken to be adjunction to IP (cf. Lasnik&Saito 1992, Chomsky 1995a, p.323), (56a) should have been saved by counter-cyclic application of topicalization, creating a "topic-island" only after WH-movement has been ruled grammatical. Likewise, the option of counter-cyclic scrambling, analyzed as adjunction to IP, incorrectly predicts (56c) to be as well-formed as (56b) (cf. Müller 1998).
We can conclude that there is reason *not* to exempt overt XP-adjunction from the extension condition on transformations (48).[105]

---

[103] Cf. Reinhart (1991). See Bolinger (1979) for a discussion of further data.
[104] Something like this seems to have been Chomsky's original line, as reported in Speas (1991, p.248), Speas discusses another type of problematic cases.
(i) [Mary's$_i$ cat] she$_i$ likes *t*
(ii) [John's$_i$ ELDEST son] he$_i$ sent *t* to camp
Possessors, standardly located in specifier positions, do not violate principle C either, although they shouldn't be inserted counter-cyclically. However, a proper theory of focus, such as the one presented in Schwarzschild (1999), could assign the NPs *Mary* and *John* the status [+given]. This would make the cases in (i) and (ii) amenable to an analysis in terms of discourse coreference rather than binding theory.
[105] Kitahara (1995) develops a system to derive the original contrast in (51) on the basis of economy principles. According to that theory, cyclic introduction of complements or specifiers requires fewer operations than its counter-cyclic variant. For adjunction, on the other hand, both variants take the same number of steps, such that optionality arises. Although that theory covers an

What we have seen so far indicates that there is no empirical reason to require more than trivial ordering of operations. This would then be virtually indistinguishable from determining well-formedness on the basis of the output representations.[106] We are only left with theory-internal exceptions to condition (48) in the domain of X°-movement and covert movement. The analysis of the latter (in sections 2.6 and 2.7) will uncover further unwelcome effects. So, I will ultimately defend the following hypothesis 3.

Hypothesis 3: There are no counter-cyclic syntactic operations (=H3)[107]

An implementation of that hypothesis will have to be worked out in section 3, where I'll also discuss X°-movement.[108]

2.6 Minimalist Operations II

Section 2.5 illustrated the broader technical consequences forced on a minimalist grammar by virtue of its interface-oriented elimination of intermediate D- and S-structure. The next step, documented in the present section, is to find out to what extent objects and operations hitherto employed in generative syntax can be fine-tuned further. This aim has been set on the agenda by Chomsky (1995a chapter 4, 1995b) under the heading "bare phrase structure."

2.6.1 Lexical Items, Numerations, and Inclusiveness

We have already seen in section 2.5.1 that X-bar schemata, the residue of former phrase structure rules, acquired a somewhat precarious status as conditions on operations, not quite in the spirit of bottom-up generation of syntactic objects. What's more, they reimport unresolved problems into the system, which I summarized as open questions under Q2.

---

interesting range of facts, I have to reject it on both empirical grounds, as have just been offered, and because of my general objection to economy principles (cf. section 2.8). Likewise, (51) shouldn't be taken as evidence for a theory of "sideward movement," although this was proposed by Nunes (1995, p.81ff).

[106] Cf. Brody (1995, 1998). For discussion see also Freidin (1978), who showed how to trivialize the then current weaker condition on cyclicity, i.e. the Strict Cycle Condition.

[107] In other words, there is a "principle of syntactic compositionality," as Peter Staudacher (p.c.) suggested. Bobaljik (1995a), Brody (1998), and Groat & O'Neil (1996) also arrive at this claim. H3 is in conflict with Rooryk's (1994) account of free relatives, which relies on counter-cyclic GT.

[108] See section 3.4.2. The exact formulation of the ban on countercyclicity is dependent on the syntactic system into which it is embedded. The proposal in section 3.3 will appeal to (i) below.

(i) Strictest Cycle Condition
Every syntactic operation creates a root node

Now, in order to streamline the system, Chomsky (1995a) relies on the minimalist strategy, discussed in section 2.2 above, to induce syntactic properties from the outside as much as possible. Nothing should be postulated that either (i) isn't needed at the sound/meaning interfaces, as has been expressed by the Principle of Full Interpretation, or (ii) isn't there in the vocabulary which syntax operates on. The latter maxim has been termed “inclusiveness.“

(57) The Principle of Inclusiveness
Any structure formed by the computation (in particular, π and λ) is constituted of elements already present in the lexical items selected for N; no new objects are added in the course of computation apart from rearrangements of lexical properties (Chomsky 1995a, p.228).[109]

Departures from inclusiveness, seen from this perspective as “imperfections“ of I-language, will have to be firmly motivated.

For the following, it is useful to consider the vocabulary (LEX) to be the lexicon closed under a number of specification operations, such as determining case on nouns and agreement on nouns and verbs. The lexical items drawn from LEX are each feature-structure triples, comprising (*morpho-*)*phonological*, *formal*, and *semantic* features, which I will call *FPHON*, *FFORM* and *FSEM* respectively. Nothing else should be available to C_{HL} according to inclusiveness. Thus, “in particular, no indices, bar levels in the sense of X-bar theory, etc.“ (Chomsky 1995a, p.228), which means among other things that the devices determining binding relationships and chain membership in GB-theory are off limits. Raising the stakes even higher, the following comment is added.

> “Note that considerations of this nature can be invoked only within a fairly disciplined minimalist approach. Thus, with sufficiently rich formal devices (say, set theory), counterparts to any object (nodes, bars, indices, etc.) can readily be constructed from features. There is no essential difference, then, between admitting new kinds of objects and allowing richer use of formal devices; we assume that these (basically equivalent) options are permitted only when forced by empirical properties of language“ (Chomsky 1995a, p.381fn.7).

Indices in particular seem to be in need of reconsideration.

> “A theoretical apparatus that takes indices seriously as entities, allowing them to figure in operations (percolation, matching, etc.), is questionable on more general grounds. Indices are basically the expression of a relationship, not entities in

[109] I'll come to “numerations,“ i.e. *N* in (57) shortly. Austere as the formulation of inclusiveness is, it potentially diverts attention from the nature of operations (e.g. copying involved in Move) that are as much constitutive of a structure as the set of objects manipulated. See section 2.3 above and hypothesis H1 of section 1.1. Note also that inclusiveness is violated at PF (cf. Chomsky 2000, p.100).

their own right. They should be replaceable without loss by a structural account of the relation they annotate" (Chomsky 1995a, p.217fn.53).[110]

Strengthening restrictions even further, only formal features should be accessible to the computational procedures of C_{HL}, were it not for the fact that PF, crucially operating on morphophonological features, is considered part of C_{HL}.[111]

Now, the point of departure for a syntactic computation is determined by a choice of lexical items from LEX, collected in a data structure called "numeration" ($\mathbb{N}$).

(58) A *numeration* is "a set of pairs (LI, i), where LI is an item of the lexicon and i is its index, understood to be the number of times that LI is selected" (Chomsky 1995a, p.225).

Given what was just said about indices, invoking them this early might come as a surprise. However, something has to be done about individuation of tokens of lexical items in syntax (cf. Chomsky 1995a, p.227).[112] For obvious reasons, the numeration couldn't be just a set, or else simple structures like (59a) and the mind-bending, unboundedly elaborable versions of (59b) from Stabler (1992, p.87) were impossible to capture, although these definitely fall into the domain of competence grammar, as opposed to a theory of performance.

(59) a. She thought she had told him
b. Police police police
c. Police police police police
d. Police police police police police
e. . . .

---

[110] In the light of what is still to come (see especially section 3.3.3), I would like to point out that in keeping with neurophysiological findings, models are being developed that rely on neural firing activity to bring about dynamic binding, synchronization of temporal firing patterns being interpretable as "coindexation" (Shastri&Ajjanagadde 1993, Bienenstock&Geman 1995). Thanks to Reinhard Blutner (p.c.) for pointing this out to me. Thus, insisting on any particular construal of indices, such as one from which it follows that they cannot be entities, seems to imply a degree of formal commitment not normally considered justifiable. See section 1.1 above. See also Chomsky (1965, p.210fn.4) where arguments are offered against Gilbert Harman's proposal to regard category symbols as "pairs (α,φ) where α is a category symbol and φ is a set of indices used to code transformations, contextual restrictions, etc.)." A similar appeal to "indexed grammars" has more recently been made by Gazdar et al. (1985). See Partee et al. (1993) and Wartena (2000) for the formal background. See also Rogers (1998), for a complexity result concerning "free indexation" as employed in GB binding theory.

[111] PF is also an exception to inclusiveness, as noted by Chomsky (2000, p.100). See also section 2.2 for a comment on the implications of the SYN/PHON-split for the conception of interfaces and "uniformity of computation."

[112] See Kitahara (2000), for an alternative approach. The question of individuation plays a crucial role in motivating use of indices in section 3.3 below.

As a first approximation, a numeration for cases like (59b) could look like the one in (60).

(60) $\mathbb{N}$ = { < <FPHON = /police/, FFORM = { < Cat, N >, < Case, Nom >, . . .},
FSEM = . . . >, 1 >,
< <FPHON = /police/, FFORM = { < Cat, V >, . . .}, FSEM = . . . >, 1 >,
< <FPHON = /police/, FFORM = { < Cat, N >, < Case, Acc >, . . .},
FSEM = . . . >, 1 > }

Lexical insertion, called "Select" (cf. 2.6.2), takes one lexical item out of $\mathbb{N}$ and reduces its index by 1. As a minimal condition on convergence the numeration must be emptied, in the sense that all indices have ultimately to be at 0 (cf. Chomsky 1995a, p.225). Yet, further adjustments are required in the light of the following assumption about syntactic operations (Chomsky 1995a, p.230).

(61) Overt operations cannot detect phonological features at all.

The same one might expect to hold for semantic features. Thus, the content of FPHON can only be accessed at the level of PF, the content of FSEM only at or beyond the C-I interface (see section 2.2).[113]
Inside the numeration, this would lead to many more indistinguishable objects, to the detriment of considering $\mathbb{N}$ a set. For example, $\mathbb{N}$ of (62) could be taken to contain a single member only.[114]

[113] This is what could be inferred from the following remark: "Among the features that appear in lexical entries, we distinguish further between *formal* features that are accessible in the course of the computation and others that are not: thus, between the formal features [+/- N] and [+/- plural], and the semantic feature [artifact]" (Chomsky 1995a, p.230). However, the reasoning is more complicated. On the assumption that ouput conditions determine what is accessible to C_{HL}, Chomsky (1995a, p.242) concludes that, since at the LF interface the nonphonological properties of lexical items, i.e. the content of FFORM and FSEM must be accessed, it must be available for C_{HL} generally. In the light of (61), this again highlights the different relations of LF and PF to syntax in minimalism, the one directly, the other indirectly linked with syntactic computations. It may be conjectured that a "hidden interface" (presumably at Spell-Out) prevents PF from direct influence on syntax, that is, if one sticks to the logic of the model in the first place. See sections 2.2 and 3.1 for an alternative. Anyway, the content of FSEM could potentially be functioning within C_{HL}. It remains to be added that, for technical reasons, encountering the attributes FPHON and FSEM at the "wrong" interface representations, i.e. λ and π respectively, should lead to FI violations. This prevents phonologically empty categories, represented as <FPHON = ∅, FFORM, FSEM> (not <FFORM, FSEM>) from being inserted into syntactic derivations after Spell-Out.

[114] In accordance with what was said in the previous footnote, the possibility can't be ruled out that syntactic tokens are individuated by the content of FSEM. Thus, the features [+ feline], [+ canine], and [+ ranine (?)] could prevent the individuation problems discussed here, though not the ones arising from examples like (59). I continue, however, to disregard this employment of features from FSEM. The idea that syntax should be kept free of morphophonological and semantic features in

(62) $\mathbb{N}$ = { < <FPHON = / cat /, FFORM = { < Cat, N >, < Case, Nom > ,
< Agr, 3rd-sg > }, FSEM = . . . >, 1 >,
< <FPHON = / dog /, FFORM = { < Cat, N >, < Case, Nom > ,
< Agr, 3rd-sg > }, FSEM = . . . >, 1 >,
< <FPHON = / frog /, FFORM = { < Cat, N >, < Case, Nom > ,
< Agr, 3rd-sg > }, FSEM = . . . >, 1 > }

The common and useful convention to present syntactic terminals in full (graphemic) spelling, which I will continue to employ for expository purposes whenever nothing hinges on this, is thus misleading from a minimalist point of view.[115] One way to remedy the situation is to appeal to unordered lists (Bobaljik 1995a), or multisets (Kracht 1997), as the appropriate data structures from which to start computations. Thus, $\mathbb{N}$ of (59b) would look like (63), if we consider only category symbols for the sake of brevity. (Functional heads have been added.)[116]

(63) $\mathbb{N}$ = { N, N, V, AgrO, T, AgrS $\}_M$

Given that a multiset over a set S can also be seen as a function, f: $S \rightarrow \omega$, such that f(s) denotes how many times s has been chosen for each $s \in S$ (Kracht 1997), the numeration could be kept in its original form as long as the lexicon is collapsed into a set of (distinguishable) FFORMs, that is, fully specified “preterminals.“ In this case, (63) would, again restricting ourselves to categorial features, be replaced by (64).

(64) $\mathbb{N}$ = { < N, 2 >, < V, 1 >, < AgrO, 1 >, < T, 1 >, < AgrS, 1 >, < Adv, 0 >,
< P, 0 >, < A, 0 >, . . . }[117]

Interestingly, such a move basically means adopting Jackendoff's (1997) position, according to which lexical insertion, in the classical sense of operating on members from the *terminal* (V_T) as opposed to the *non-terminal* vocabulary (V_{NT}), takes place only after syntactic computations are completed. This would further require to have PF branch off from λ, which contradicts the standard minimalist view that Spell-Out is possible before the computation has reached λ (cf. (24b) of section 2.4). I'm going to advocate the corresponding modification in section 3 on independent grounds.[118]

the first place has recently been advocated by Jackendoff (1997). See also Halle&Marantz (1993) for a related system as far as morphophonology goes.

[115] See Pollard&Sag (1987, 1994) for a representational format that properly captures the “marginality“ of this spelling for syntactic theory.

[116] Subscripted M signals use of multisets (cf. Kracht 1997). See also Gazdar et al. (1985) for the employment of multisets in defining ID-rules.

[117] If partial functions are assumed, none of the zeroed elements have to appear. Given that numerical devices have been criticized as “too strong“ for human language, e.g. in the area of X-bar levels (Zwarts 1992, p.15), no strong commitment to the particular view of numerations in Chomsky (1995a) is to be expected.

[118] Cf. Brody (1995, 1998) and Groat&O'Neil (1996). See also section 2.2 above and 2.7 below.

2.6.2 Select, Merge, and Move

Next, syntactic derivations have to be defined as transitions from the resources collected in ℕ to the designated representations π and λ (cf. Chomsky 1995a, p.225f).

(65) A *derivation* is a sequence of stages $\Sigma_0, \ldots, \Sigma_n$ such that for each i ($0 < i \leq n$), Σ_i is the outcome of exactly one syntactic operation applied to Σ_{i-1}.

(66) A *derivational stage* Σ is a set of syntactic objects ($\Sigma = \{SO_1, \ldots, SO_n\}$)

The most elementary syntactic operation is "Select."

(67) *Select* removes a lexical item from the numeration, (reduces its index by 1), and introduces it into the derivation as SO_{n+1} (Chomsky 1995a, p.226).

Again, if derivational stages are sets, indices have to be used unless one somehow constrains the order of insertion into, and other operations on Σ, such that indistinguishable syntactic objects never have to cooccur. Taking Σ to be a multiset, as proposed for ℕ in section 2.6.1, would obviate the need for any such extra assumptions. The conservation of objects would be warranted and operations can apply in any order.[119]

Note that derivational stages are individuated by the syntactic objects they contain. These serve as source objects for the transformational mappings. Thus, carried over from section 2.5.1, we get generalized transformations, now called "Merge," which "takes a pair of syntactic objects (SO_i, SO_j) and replaces them by a new combined syntactic object SO_{ij}" (Chomsky 1995a, p.226). Singulary transformations, called "Move," apply to a single syntactic object. As before, this crucially involves the copying of a part of SO, the details of which will be discussed below.

Let's look at a very much simplified sample derivation. For technical reasons, I take a derivational stage Σ to be an ordered pair of a set of syntactic objects and a numeration (cf. Collins 1997, p.3).

(68) a. Σ_0 = < ∅, ℕ = { < Kermit, 1 >, < likes, 1 >, < beans, 1 > } >
b. Select
c. Σ_1 = < { SO_1 = likes }, ℕ = { < Kermit, 1 >, < likes, 0 >, < beans, 1 > } >
d. Select
e. Σ_2 = < { SO_1 = likes, SO_2 = beans }, ℕ = { < Kermit, 1 >, < likes, 0 >, < beans, 0 > } >
f. Merge

[119] It may be asked, whether a data structure like Σ is required over and above ℕ. Thus, Kracht (1997) proposes to define operations directly on ℕ, rendering Σ and Select superfluous. Bobaljik (1995a) and Collins (1997) advocate the elimination of numerations instead. As far as more than terminology is at stake, this is likely to have to do with economy principles, for which the numeration is supposed to provide the "reference set" (cf. Sternefeld 1997). See section 2.8 below.

g. Σ_3 = < { $SO_{1/2}$ = { likes, beans } }, $\mathbb{N}$ = { < Kermit, 1 >, < likes, 0 >,
< beans, 0 > } >
h. Select
i. Σ_4 = < { $SO_{1/2}$ = { likes, beans }, SO_3 = Kermit }, $\mathbb{N}$ = { < Kermit, 0 >,
< likes, 0 >, < beans, 0 > } >
j. Merge
k. Σ_5 = < { $SO_{1/2/3}$ = { Kermit, { likes, beans } } }, $\mathbb{N}$ = { < Kermit, 0 >,
< likes, 0 >, < beans, 0 > } >
l. Move
m. Σ_6 = < { $SO_{1/2/3}$ = { beans, { Kermit, { likes, ($_{\text{COPY}}$ beans) } } } },
$\mathbb{N}$ = { < Kermit, 0 >, < likes, 0 >, < beans, 0 > } >

By applying Spell-Out (cf. section 2.4) to $SO_{1/2/3}$ of Σ_6, (68) ultimately derives the string (69) at π.[120]

(69) Beans, Kermit likes

Note that nothing much has changed in comparison with derivation (42) in section 2.5, except for the apparent absence of projection operations. However, a closer look at the definitions reveals the underlying drift toward eliminating superfluous entities. Take Merge first.

(70) *Merge* [is a binary, commutative, and non-associative operation that]
a. applies to two (syntactic) objects α and β [of Σ_i], and
b. forms the new (syntactic) object { γ, { α, β } } [of Σ_{i+1}] such that γ, called the "label" of { γ, { α, β } }, is constructed from the head of either α or β, and
c. eliminates α and β [$\alpha \notin \Sigma_{i+1}$ and $\beta \notin \Sigma_{i+1}$].[121]

Merge is supposed to be one of the elementary steps (cf. section 1.1), forced on minimalist grammar by "virtual conceptual necessity." Thus, set formation is taken to be the simplest binary operation adequate for imposing constituent structure on lexical items. First of all, it is commutative, that is, β ¤ (α ¤ { }) = α ¤ (β ¤ { }), as opposed to concatenation ($\alpha + \beta \neq \beta + \alpha$).[122]

---

[120] Given our discussion in section 2.6.1, things could look simpler. $\mathbb{N}$ = { N, V, N $\}_M$ or $\mathbb{N}$ = { Kermit, likes, beans $\}_M$ would be transformed into $\mathbb{N}$ = { $\}_M$ and no indices have to be attached to syntactic objects.

[121] This definition fuses the relevant passages from Chomsky (1995a, p.226 and p.243ff). Bracketed parts are my own additions (H.M.G.). Parts in parenthesis indicate where formulations differ. See section 2.6.3 for a closer look at labels.

[122] Cf. Manna&Waldinger (1985, p.473). ¤ is an insertion operator. Commutativity, in contrast to non-commutativity, is a somewhat odd property, as has for example been noted by Quine (1995/1941[1], p.5) mentioning "monograms," i.e. the graphical superimposition of two symbols, as a way of avoiding "excess of notation over subject-matter."

Such a turn from string-based to set-based syntax had been discussed in Chomsky (1965, p.124ff), where proposals in that direction by Curry (1961) and Šaumjan& Soboleva (reviewed by Hall 1964) were dismissed, among other things on the grounds that the linear order of elements, relevant at PF, requires the set system to be converted into a concatenation system by some extra device (Cf. Speas 1990).

> “In fact, no proponent of a set-system has given any indication of how the abstract underlying unordered structures are converted into actual strings with surface structures“ (Chomsky 1965, p.125).

This problem, however, can now in principle be solved on the basis of Kayne's (1994) “Linear Correspondence Axiom“ (LCA), inducing the precedence relation among terminals from asymmetric C-Command among non-terminals. In minimalist syntax, such a linearization operation is delegated to the PF component, given that

> “[t]here is no clear evidence that order plays a role at LF or in the computation from N to LF“ (Chomsky 1995a, p.334).

Syntax, i.e. C_{HL} minus PF, *can*, and therefore, according to the logic of minimalism, *must* operate on sets instead of strings. I'll come to the precedence relation in section 3.[123]

Turning to associativity, we have to note that Merge differs from both concatenation and set insertion. Thus, while $(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)$ and γ ¤ (β ¤ (α ¤ { })) = α ¤ (β ¤ (γ ¤ { })), the narrowly syntactic part of C_{HL} is taken to disallow so called “flat“ structures, in favor of strict binary branching. This property is guaranteed by (i) the binary nature of Merge, creating a new set from two objects, and (ii) by the fact that Merge is a set-constructor (cf. Kracht 1997) as opposed to operations like insertion or union, the latter alternatives, of course, being linguistically inadequate for treating lexical items as atomic in the way it is done here.[124]

By restricting the domain of the set-constructor Merge to *syntactic objects*, i.e. members of Σ, cyclicity, as expressed by the Extension Condition on Transformations (cf. 2.5.2), automatically follows for this type of generalized transformation. In other words, Merge, seen from the perspective of trees, must result in the addition of a root node immediately dominating the roots of the input trees (cf. Chomsky 1995a, p.248). Thus, to recap our earlier discussion, trying to derive (51b) one would arrive at the following derivational stage Σ in (71).

(71) Σ = { SO_1 = $[_{CP}\ [_{DP}$ Which claim $]_j$ was he_i willing to discuss t_j],
SO_2 = $[_{CP}$ that $John_i$ made] }

---

[123] See Bayer (1996) for discussion of “directionality and logical form.“

[124] Binarity would follow from an across the board implementation of Kayne's LCA in syntax, except for an additional need of non-branching projection. Collins (1997, p.75ff) suggests that binarity can be derived from his “Minimality Condition“ on a par with constraints on movement. See Brody (1998), for a system that allows ternary branching, eliminating intermediate projections.

However, application of Merge at stage Σ cannot produce Σ' in (72a), the required structure for relative clause adjunction. That is because *DP* (= *which claim*) of SO_1 is not itself a syntactic object, so it is not in the domain of Merge. Instead, the only possible output of Merge applied at Σ is Σ'' in (72b), a linguistically dubious result as far as constituency and word order are concerned.[125]

(72) a. * Σ' = { $SO_{1/2}$ = [$_{CP}$ [$_{DP}$ [$_{DP}$ Which claim]$_j$ [$_{CP}$ that John$_i$ made]] was he$_i$ willing to discuss t$_j$] }
b. Σ'' = { $SO_{1/2}$ = [$_{CP}$ [$_{CP}$ [$_{DP}$ Which claim]$_j$ was he$_i$ willing to discuss t$_j$] [$_{CP}$ that John$_i$ made]] }

This, then, goes half-way toward my hypothesis 3 (cf. section 2.5.2), banning counter-cyclic operations in general. The technical means for analyzing in a counter-cyclic fashion the *prima facie* empirically most compelling data in favor of countercyclicity have been removed.[126]

Now, further fine-tuning is called for as soon as the singulary transformation Move is considered. As illustrated in (68k)-(68m), Move transforms a single syntactic object of a derivational stage Σ, keeping its status as a syntactic object constant. The internal workings of Move, however, are more complex. First, one needs to identify two elements, the target and the item that “moves.“ The latter I will also sometimes call “voyager.“ Then, there has to be a copying operation in order for there to be a trace left in the position from which something moves. Also, there is an operation that looks like Merge, applying to the voyager and the target. Further, we get chain formation (cf. 2.5.1).

From the perspective of derivations, Move seems to affect a different layer of objects inside the hitherto recognized source objects called “syntactic objects.“ Recall that Merge eliminates its input elements from the set of syntactic objects. Thus, *beans* in (68) is a syntactic object at stage Σ_2. The input to Move is Σ_5, though, which does no longer recognize *beans* as a syntactic object. The extra layer of objects required, is consequently provided by the introduction of “terms.“[127]

---

[125] Recall that the output of Merge is unordered. The only reasonable ordering of (72b), however, would have the relative clause in rightwardly extraposed position, which, although well-formed in English, is not obligatory of course. Note also that neither (72a) nor (72b) is compatible with Kayne's (1994) LCA, so the minimalist PF-implementation of that principle seems to rule these structures out as well.

[126] In reconsidering a richer definition of Merge that would allow counter-cyclicity, Chomsky (1995a, p.248) notes: “Any such complication (which could be quite serious) would require strong empirical motivation. I know of none, and therefore assume that there is no such operation. Merge always applies in the simplest possible form: at the root.“ But see Chomsky (1995a, p.327) for a caveat wrt adjunction.

[127] Note the relational character of terms, marking them as secondary objects. As far as I can see, replacing the rather indeterminate reference to “structure“ in (73) by “syntactic object“ would be a feasible modification. Such a change is tacitly assumed by Nunes (1995, p.64) and Nunes&Thompson (1998). This would for example prevent the definition of terms from applying to derivational stages, which could produce unwelcome results. Since the system in section 3 does

(73) For any structure K,
a. K is a *term of* K.
b. If L is a term of K, then the members of the members of L are *terms of* K.

A prerequisite for the definition of terms to be adequate is to regard the members of $\mathbb{N}$ as atomic in the sense that they are not sets and thus do not have members (cf. section 2.6.3). Viewing them as <FPHON,FFORM,FSEM>-triples may be taken to meet this condition. As an effect, terms can be put in a one-to-one correspondence with nodes or subtrees of constituent structure trees. Let's look at (74), a still simplified close-up of $SO_{1/2/3}$ of Σ_5 in (68k).

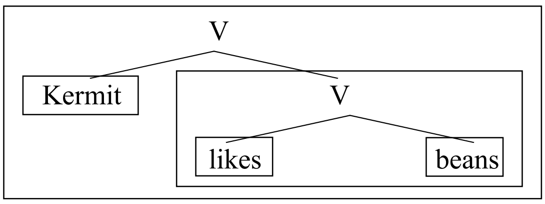
(74) $SO_{1/2/3}$ = { V, { Kermit, { V, { likes, beans } } } }

Applying definition (73) to (74), we arrive at the five terms of (75).

(75) a. { V, { Kermit, { V, { likes, beans } } } }
b. Kermit
c. { V, { likes, beans } }
d. likes
e. beans

Mapped into a tree structure, (75) requires five nodes, each the root of a unique subtree. This is shown in (76). (Subtrees are enclosed in boxes.)

(76)



Thus, what I've called "node record" of GT and Move-α in 2.5.1 can be carried over as a "term record" of Merge and Move. The ambiguous Project-α, which has no independent status any longer, is replaced by "trivial" Select. The latter can be taken to convert an item that does not count as a term, i.e. a member of $\mathbb{N}$, into an item that does, i.e. a member of Σ.

(77) Term Record
a. Select (0) → 1
b. Merge (n,m) → n+m+1 ($n > 0$, $m > 0$)
c. Move (n) → n+m+1 ($n > 0$, $m \leq n$)

---

away with the distinction of syntactic objects and terms, the counterparts of the latter will not have to be defined relationally.

Still, it has to be remembered that, in line with the principle of inclusiveness, trees and nodes do not have any status within minimalist syntax. Whatever notion is deemed indispensable has to be reconstructed.[128]

Movement, to begin with, clearly requires the displacement not just of arbitrary (collections of) nodes but of subtrees, although for borderline cases the difference may not be detectable. So if terms are taken to provide the appropriate objects for the Move operation, terms should exhibit properties of subtrees.

> "Terms correspond to nodes of the informal representations, where each node is understood to stand for the sub-tree of which it is the root" (Chomsky 1995b, p.399).[129]

Indeed, as far as notation goes, it is much easier to see that terms "stand for" subtrees, since the entire set of objects and relations is given in that notation (cf. (75)).[130]
Now, it looks as if the singulary transformation Move could be defined as follows.

(78) *Move* [preliminary][131]
a. applies to a syntactic object SO_i of Σ_j
b. identifies two terms τ_1 and τ_2 of SO_i, τ_1 called "target" and τ_2 called "voyager"
c. produces τ_2', a copy of τ_2
d. merges τ_2' and τ_1 to produce $\tau_3 = \{ \gamma, \{ \tau_2', \tau_1 \} \}$
e. replaces τ_3 for τ_1 in SO_i of Σ_{j+1}
f. forms a chain, CH = $< \tau_2', \tau_2 >$
g. marks τ_2 as invisible at π[132]

[128] Obviating a lot of complications, Stabler (1996, 1998) replaces the set-based syntactic objects by ordered trees without much ado. Thus, to the extent that Stabler's acquisition and parsing results are taken as vindications of minimalism, it must be kept in mind that these are gained on the basis of linguistically well-understood objects, the congruence of which with bare phrase structure still requiring closer examination.

[129] In the context of minimalist syntax, objects are considered "informal," it seems, i.e. used in informal exposition, if they aren't defined in minimalist syntax. In Chomsky (1995a, p.247), which was written later than Chomsky (1995b), taking the latter text as starting point, this quote is qualified by adding "for the case of substitution." Although this is clearly meant to filter out adjunction structures, it is unclear why adjunction structures should differ along the node/subtree dimension (cf. section 2.6.3). See sections 3.3.1 and 3.3.3, where the subtree/node distinction gives rise to two distinct systems.

[130] This perspective seems to be further confirmed by the following assertion: "Interpreting the informal notation more closely, we might say that *the tree with root* T' is {T,{T,K}} (etc.), to overcome the temptation to think of T' as having some distinct status, as it does in phrase structure grammar (and its X-bar theoretic variant)" (Chomsky 1995a, p.383fn.30).

[131] See (80) below for the published version.

[132] As already mentioned, Move is usually said to leave a copy in the original position of the "voyager," so that the procedure just outlined is further complicated in that τ_2 and τ_2' have to change places after the copy operation in (78c). Further conditions on the nature of the voyager and

Given (78d), one might be inclined to take Merge as an algorithmic module of Move (cf. section 2.5.2). Such an assumption has been made in alternative approaches, where Move is considered epiphenomenal in the sense that its properties follow from independent conditions on the suboperations Copy, Merge, Form Chain, and Delete.[133]

Bobaljik (1995a, p.41), whose alternative approach closely resembles what I will propose in section 3, regards the coexistence of Merge and Move as "redundant." Little attention, however, is paid to the "ontology" of the theory.[134] Thus, if Merge is defined for *syntactic objects* in (70), it cannot be a module of Move, since here Merge has to apply to *terms*, the striking similarity of outputs notwithstanding.

At this point it is recommendable to have another look at the definition of Merge, repeated as (79) for convenience.

(79) *Merge* [is a binary, commutative, and non-associative operation that]
a. applies to two (syntactic) objects α and β [of Σ_i], and
b. forms the new (syntactic) object { γ, { α, β } } [of Σ_{i+1}] such that γ, called the "label" of { γ, { α, β } }, is constructed from the head of either α or β, and
c. eliminates α and β [$\alpha \notin \Sigma_{i+1}$ and $\beta \notin \Sigma_{i+1}$].

In fact, the domain of Merge has not really been fixed, due to the existence of two alternative formulations. One way to proceed would be to give up the restriction to *syntactic* objects. Merge could then apply to terms as well.

Then, of course, one has to specify what "objects" are in order to preclude unwelcome consequences. Secondly, elimination of α and β in (79c) only makes sense on the level of syntactic objects, which constitute the source objects for singulary and binary transformations by providing the members of derivational stages. Transformations then proceed in the way illustrated by transformation markers in Chomsky (1965, p.130). On the level of terms, α and β remain constant. Thirdly, what has been said about cyclicity wrt Merge becomes unclear. Given that in (71) *DP* (= *which claim*) of SO_1 is a term, SO_2 could be merged with that DP, so the possibility of producing (72a) could no longer be ruled out without extra assumptions.[135]

Now, in the light of the vagueness accompanying the domain of Merge, it would be risky to modify the definition of *terms* (73) by simply replacing reference to *structure* by reference to *syntactic objects*. I would rather raise the more fundamental issue as to whether the "ontology" implied by the system so far squares well with the strict

the target, deriving from checking theory and locality, are to be integrated into the definition of Move. Recall the two heuristics from section 1.1. The complexity of Move is certainly motivated from beyond the borderline for elementary constructive steps. Clearly, it is determined by intended output, namely, the capturing of (unbounded) dependencies.

133 Cf. Bobaljik (1995a), Nunes (1995), and Collins (1997). See also Chomsky (2000, 2001). The same line is taken from an even more abstract point of view by Koster (1987) and Brody (1995, 1998).

134 I'm using the term "ontology" in an extended, less rigid, sense here. Thanks to Günther Grewendorf (p.c.) for pointing this out to me.

135 See Collins (1997, p.81ff) for one solution.

dichotomy of singulary and generalized transformations customary for minimalist syntax. More precisely, two closely related questions arise.

Q3: (i) Is there any need for both syntactic objects and terms?
(ii) Should there be more than one type of Merge?

The answer given to Q3(ii) in section 3 is negative, insofar as Merge and Move are going to be compounded into a single, hybrid operation. Likewise the answer to Q3(i) is negative, given that the irreducible differences between Merge and Move can be located elsewhere in the theory.[136]

Note that what has been said so far about the complications of integrating Move into the system is based on my preliminary conjecture, expressed in definition (78), that Move crucially applies to syntactic objects and involves a Merge operation defined on terms. This, however, is not a fully accurate rendition of Chomsky (1995a). Neither Merge nor syntactic objects are directly appealed to in the crucial passage. Thus, the characterization of Move is the following (Chomsky 1995a, p.250).[137]

(80) Move
"Suppose we have the category Σ with terms K and α. Then we may form Σ' by *raising α* to *target K*. That operation replaces K in Σ by L = { γ, { α, K } }."

Likewise, both Move and Merge are occasionally called *operations that form categories* (e.g. Chomsky 1995a, p.250). Thus, talk of categories has to be harmonized with talk of terms and syntactic objects, over and above the task of clarifying the open questions attached to the latter notions.

In sum, we have some grasp of how Merge and Move operate, e.g. (68), although characterizations like (70), (78), and (80) still display a considerable amount of vacillation.[138]

---

[136] See also section 3.5, for some additional consideration of these differences. In order to deal with Move, both Nunes (1995) and Collins (1997) allow for the possibility of copying *terms* from *syntactic objects* and treating them as *syntactic objects* of a later derivational stage, which amounts to a kind of "type-lifting." Collins (1997) avoids reference to "terms," speaking of "constituents" instead.

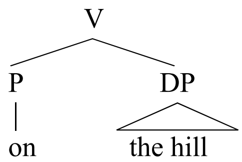
[137] The precursor of Chomsky (1995a) still explicitly contained reference to Merge in characterizing Move (Chomsky 1995b, p.399). This also made its way into some formulations of chapter 4 (cf. Chomsky 1995a, p.250), for one example. Consider also the following most explicit case. "It [C_{HL}] consists of two operations, Merge and Attract/Move, which incorporates Merge" (Chomsky 1995a, p.378).

[138] This is a somewhat irritating property of the minimalist program. Although great efforts are made to present an optimal design for the minimal of all possible systems, terminology is definitely not minimal. Whereas certain notions like trees, nodes, indices, and bar-levels are explicitly regarded as undefined within minimalism, even as "informal," it is not quite clear which notions of minimalist grammar are really defined, not to speak of formal. See Chomsky (2000, 2001) for some clarifications.

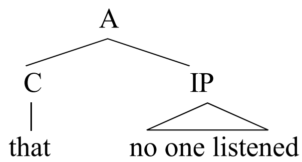
2.6.3 Labels, Projections, Categories, and Adjunction

Developing a theory of bare phrase structure, as discussed in sections 2.6.1 and 2.6.2, implies that no recourse to X-bar schemata (section 2.5.1) is allowed. Viewing syntax as defining a recursive structure of sets on a collection of members from LEX can certainly reconstruct the concatenative and hierarchical aspects inherent in phrase structure rules. But what about the categorial information transported by linguistic expressions? One of the tasks of X-bar rules was to preclude ill-formed categorizations, such as the ones in (81).

(81) a. *



b. *



Recall, however, that in keeping with inclusiveness, syntax is lexically driven. Moreover, categorial information is located inside FFORM in the guise of categorial features, e.g. <CAT,N>, or equivalently <CAT,<<N,+>,<V,->>>.[139] At the same time, there is no independently recognized non-terminal vocabulary, V_{NT}, which could undergo concatenation or provide values for a labeling function (cf. section 2.3), if such a function were assumed.

Now, the minimal way to rule out ill-formed categorizations like (81) would be not to categorize at all. Why not freely combine (sets of) members from LEX?[140] Chomsky (1995a) pursues a more traditional way, letting each operation of Merge non-deterministically designate one of its arguments as providing the type of the operation's output.[141] Thus, strictly speaking, Merge is a recursive set-constructor, forming $\{ \alpha, \beta \}$ from α and β first and subsequently $\{ \gamma, \{ \alpha, \beta \} \}$ from $\{ \alpha, \beta \}$ and the "label" γ, γ being the terminal head of either α or β, i.e. *th*(α) or *th*(β). For reasons of transparency, Merge could be seen as two operators <♦,♣> applying to the input in sequence.[142]

---

[139] Cf. Chomsky (1995a, p.236).

[140] Such a radical possibility, as discussed by Chomsky (1995a, p.244), would seem to require that any syntactically well-formed set $\{ \alpha, \beta \}$ uniquely determines its type. Stabler (1996, 1998) can be taken to imply such a position. The feasibility of this seems to depend on the question to which extent feature checking can be assimilated to function application (or cancelation) in categorial grammar. See sections 3.4.2 and 3.5 for discussion of whether selectional requirements should be dealt with in terms of feature checking. This relates to the (conditional) hypothesis 2 of section 2.4.

[141] Ideally, the wrong choice of argument in this procedure should then be banned by independent principles of the grammar. These principles are yet to be specified. See Chomsky (1995a, section 4.4.2), for a discussion of how to guarantee the output of Move to be determined by the target.

[142] ♦ is the trivial set-constructor introduced in section 2.6.2. Binarity, commutativity, and non-associativity still hold for Merge as a whole. From a non-derivational perspective this definition looks somewhat suspicious, given that *th*(x) itself is recursively determined. Thus, two recursive definitions have to interleave properly. In the case at hand, this seems to be harmless, since in each case the basis of recursion can be determined independently. This may be made explicit as follows.

(82) a. $\clubsuit(\blacklozenge(\alpha,\beta))$
b. $\blacklozenge(\alpha,\beta) = \{ \alpha, \beta \}$
c. $\clubsuit(\{ \alpha, \beta \}) = \{ \gamma, \{ \alpha, \beta \} \}$, for $\gamma \in \{ th(\alpha), th(\beta) \}$

The terminal heads recognized in syntax are restricted to members of LEX. For a particular derivation, the set of terminal heads is identical to $\mathbb{N}$ if we take $\mathbb{N}$ to be a multiset. Carrying over some terminology from X-bar theory, the argument of Merge that determines the label of the output can be called "projector" and the output object is a "projection" of the (terminal head of the) projector.

Given the mechanism just described, it is easy to see that cases like (81) cannot arise. Merging *P* and *DP* (= *{ D, { D, N } }*) in (81a) only licenses a label of category P or D, the respective terminal heads of the input arguments, i.e. the output can either be a projection of P or a projection of D, not, however, a projection of V, nor N for that matter, N not being the terminal head of DP. The same holds *mutatis mutandis* for (81b). Categorial "endocentricity" of syntactic objects, as guaranteed originally by X-bar theory (section 2.5.1), automatically follows.

Again, a closer look at the labeled objects obtained reveals some special properties of the set-based syntax pursued in minimalism. In fact, labels clearly belong to a tree-based formalization of syntactic structure. Thus, a set of nodes (or vertices) over which dominance and precedence relations are defined gets associated with labels, exactly one label per node (cf. (14) in section 2.3). Labels, on that view, are drawn from V_T and V_{NT}, the terminal and non-terminal vocabulary respectively, and assigned in a way that only leaves are associated with members of V_T, all other nodes with members of V_{NT}.

Now, commitment to any particular formalization, such as trees, has never been very strong in Chomskyan generative grammar.[143] One of the salient alternatives has been to treat tokens of members from V_T and V_{NT} to be the objects directly operated on by concatenation and set-formation (Chomsky 1975/1955[1]). Under the latter perspective, talk of labels doesn't suggest itself as immediately useful. So, when presenting minimalist syntax, one has to be doubly careful, because neither trees nor V_{NT} are recognized here. What is called "label," then, must, observing inclusiveness, either be a member of LEX or an object constructed from such members. Indeed, Chomsky

Assume that Merge applies to syntactic objects, i.e. members of Σ (cf. 2.6.2). Also, there has to be a property of being "elementary" in some sense. Let's assume that being a "non-set" meets that requirement, and that members of LEX, i.e. the <FPHON,FFORM,FSEM>-triples collected in $\mathbb{N}$ and targeted by Select are non-sets. Then the following definition of "terminal head" should be adequate.

(i) Terminal Head
For every syntactic object x,
a. if x is elementary, then $th(x) = x$, and
b. if x is non elementary, such that $x = \{ \gamma, \{ \alpha, \beta \} \}$, γ elementary, then $th(x) = \gamma$.

Given that Merge necessarily operates on elementary syntactic objects first, in which case the terminal head of any resulting complex syntactic objects can be determined by applying (ia) to the input arguments of Merge. But this automatically provides the elementary γ of (ib) for applications of Merge that involve non-elementary syntactic objects.

[143] Cf. Chomsky (1982, p.14f) and discussion in section 3.2 below.

explicitly assumes that the label *is* the terminal head of one of the arguments of Merge.[144] This is, of course, trivial from a set-theoretic perspective where token identity means identity. Yet, it can be asked what the relation between syntactic theory and the syntax of set-theoretic notation is. In particular, great care is taken in the theory of movement to interpret multiple tokens of an identical type as *distinct copies*, a point which I will come back to later.[145] In any case, as far as labels are concerned, "excess of notation" is apparently rendered harmless by the definition of terms (repeated here for convenience). Terms are supposed to provide the prime "functioning elements" of syntax (Chomsky 1995a, p.247).

(83) For any structure K,
a. K is a *term of* K.
b. If L is a term of K, then the members of the members of L are *terms of* K.

Thus, labels and the sets containing the arguments of the operation Merge, i.e. the members of K, will not constitute additional terms.[146]

Now, in the section 2.6.2, we couldn't fully straighten out the relations between Merge and Move and the double-layer of syntactic objects and terms. This was partly due to the fact that talk of "categories" took over at decisive points in the original exposition. Therefore it would be of some advantage to know better what categories are. An intuitively pleasing proposal would be to regard categories as sets of expressions, as indicated in (84).[147]

(84) a. DP = { the author of Waverley, Pegasus, they, a modest proposal, . . . }
b. P = { on, over, at, in, . . . }
c. . . .

However, what an expression is likewise depends on the theory one adopts, so the advance may be only apparent.[148] Unfortunately, the definition of "category" seems to be closely bound up with the kind of phrase structure theory advocated. Thus, Chomsky

---

[144] Chomsky (1995a, p.244, p.246, p.249). Other formulations allow the conclusion that the label is *identical to* the terminal head of one of the arguments of Merge (p.248).

[145] See in particular sections 2.7.1 and 3.3.2.

[146] See the term record (77) of section 2.6.2. It would be interesting to know what effect labels have at the interfaces. Obviously, if "verbal and nominal elements are interpreted differently at LF and behave differently in the phonological component" (Chomsky 1995a, p.243) and if it is the label that indicates such categorial information for complex constituents, then the label must be detectable at the interfaces. Now, what if the FFORM of the label contained uninterpretable features, assuming that checking theory operates on terms not labels. Shouldn't they violate FI? If the answer is that *identity* of label and terminal head ensures that no such features reach π and λ, this would seem to involve an appeal to the power of set theory the legitimacy of which will be discussed in section 2.7.1 and 3.3.2.

[147] Cf. Gazdar et al. (1985, p.40fn.1) and von Stechow&Sternefeld (1988, p.144).

[148] In Chomsky (1965, p.63ff), the expressions in question were (sub-)strings "each of which is assigned to a certain category."

(1965, p.120ff) clearly identified the “categorial component“ of the grammar with its set of phrase structure rules strictly separated from the lexicon and lexical insertion, the latter brought about by substitution transformations. At the same time, there was a strict separation of non-terminal vocabulary consisting of the “category symbols“ and the terminal vocabulary represented by (lexical and grammatical) “formatives“ (Chomsky 1965, p.65).[149]

Over the years, lack of uniformization created a number of slightly incongruent ways of talking about categories.[150] Most important in our context are the symbol/string and node/subtree ambiguities.[151] For minimalist syntax, I conjecture that categories (not categorial features) must be identified with what under a tree-based formalization is called subtrees. This usage can already be located in pre-minimalist formulations, according to which categories “move,“[152] a property that linguistically only makes sense for subtrees not nodes. Also, categories “consist of“ subconstituents,[153] again a property that wouldn't be adequate for nodes. The latter property is equally appealed to in Chomsky's characterization of Move, (80), speaking of a “category with terms.“ In sum, being a category is a property of exactly the kind of objects that can be terms, namely, terminal heads and objects created from them by Merge and Move.[154]

Now, there is one minor and one major complication to report on, namely, indistinguishable projections and adjunction operations. It has repeatedly been noticed that verbal clusters, recursive prepositional phrases, or CPs as specifiers of CPs, (85a)-(85c), are not properly analyzable on the basis of bare category symbols.[155]

(85) a. weil er das hat [$_V$ [$_V$ [$_V$ machen] [$_V$ können]] [$_V$ sollen]]
because he that has make can should
“because he has been obliged to be able to do that“
b. [$_P$ [$_P$ from] [$_P$ [$_P$ under] [the table]]] (Zwarts 1992, p.16)
c. [$_C$ [$_C$ Maria [$_C$ [$_C$ wußte] [es]]] [$_C$ [$_C$ behauptete] [Xenia]]]
Mary knew it claimed Xenia

---

149 See section 3.3.3 for a return to this kind of conception.

150 Defining categories has been an immediate concern of Gazdar et al. (1985, chapter 2) and Pollard& Sag (1987, chapter 3).

151 One way these ambiguities could have arisen is the following. “We might just as well eliminate the distinction of feature and category, and regard all symbols of the grammar as sets of features“ (Chomsky 1970, quoted from Webelhuth 1995, p.23). Such an assumption would allow to talk of the symbols, i.e. members of V_T and V_{NT}, as categories. Since the symbols label nodes, nodes can derivatively be called categories as well. Finally, given that nodes are taken to “stand for“ subtrees (cf. Chomsky 1995a, p.247), even subtrees can be called categories.

152 Chomsky [with Howard Lasnik] (1995a, p.43).

153 “Each lexical category X (X = N, V, A, P) heads a category X' (X-bar) consisting of X and its complements“ (Chomsky 1986a, p.160).

154 “[. . .] the second operation that forms categories: Move (Move α)“ Chomsky (1995a, p.250). The first operation, of course, being Merge. I hesitate to say more about the relation between categories and terms for reasons having to do with adjunction, to be discussed shortly. It looks, however, as if one can neither establish that every category is a term nor that every term is a category.

155 See also the discussion of “self-adjunction“ in Chomsky (1995a, p.321).

Given that only the content of FFORM is accessible in syntax, it is hard to tell which category is a projection of which. One solution suggesting itself is to devise a mechanism able to record checking or saturation relations between the relevant elements.[156]

The more substantial complication of the theory of phrase structure so far concerns the question how to treat adjunction. The intuition behind adjunction is pretty easy to grasp from semantics and its close syntactic ally, categorial grammar. Thus, both *to snore* and *to snore violently* can be seen as intransitive verbs or predicates requiring a single argument (cf. Partee 1975).

(86) a. $[_{t}\ [_{e}\ \text{Hank}\ ]\ [_{e\backslash t}\ \text{snores}\ ]\ ]$
b. $[_{t}\ [_{e}\ \text{Hank}\ ]\ [_{e\backslash t}\ [_{e\backslash t}\ \text{snores}\ ]\ [_{(e\backslash t)\backslash(e\backslash t)}\ \text{violently}\ ]\ ]\ ]$

Adding *violently* to *snores*, "preserves" the type of the latter. In generative syntax, adjunction operations, the paradigm cases of which apply to adverbial modifiers, are said not to create a new category but to somehow "extend" the category operated on. Technically, one speaks of two (or more) "segments" of a category, one segment being the object adjoined to and the other being the resulting object.[157] Allowing adjunction operations in minimalism means to define a second type of Merge operation. Since Chomsky (1995a, p.248) locates the difference between standard Merge and adjunction in the shape of the resulting label, we can use our recursive analysis of Merge above and postulate a second sequence of operations, which takes care of adjunction.[158]

(87) Merge
a. standard: $\text{Merge}_{\langle\blacklozenge,\clubsuit\rangle}(\alpha,\beta) = \clubsuit(\blacklozenge(\alpha,\beta)) = \clubsuit(\{\ \alpha,\ \beta\ \}) = \{\ \gamma,\ \{\ \alpha,\ \beta\ \}\ \}$,
for $\gamma \in \{\ th(\alpha),\ th(\beta)\ \}$
b. adjunction: $\text{Merge}_{\langle\blacklozenge,\spadesuit\rangle}(\alpha,\beta) = \spadesuit(\blacklozenge(\alpha,\beta)) = \spadesuit(\{\ \alpha,\ \beta\ \}) =$
$\{\ \langle\gamma,\gamma\rangle,\ \{\ \alpha,\ \beta\ \}\ \}$, for $\gamma \in \{\ th(\alpha),\ th(\beta)\ \}$

I have been vague about what kind of object a "segment" should be. The most perspicuous answer would be to say that segments are *terms*. Thus, we can explicate the special nature of adjunction operations on the basis of our term record, by contrasting it with a newly invented "category record." Of course, for standard Merge, the category record yields values identical to the ones of the term record. For Merge by adjunction, the term record behaves as usual, which should be verifiable from the definition of

---

[156] (85a) and (85b) probably allow insertion of functional projections, such that an alternation of lexical and functional projections results. This may obviate the problem of finding out which projection belongs to which projector.

[157] Chomsky (1995a, p.322) speaks of two categories, however. In standard X-bar theory, adjunction is captured by keeping the bar-level of the resulting object constant, i.e. identical to the bar-level of the object adjoined to. See rules (43a), (43b), (43g), and (43h) of section 2.5.1 and see section 2.5.2 for various linguistic examples of adjunction.

[158] Assuming non-elementary labels of type $\langle\gamma,\gamma\rangle$ jeopardizes the recursive determination of terminal heads.

terms in (73), i.e. one term is added.[159] The category record, however, falls behind in that no new category is added. Rather, two terms jointly define one category.

(88) Category Record[160]
a. Select (0) → 1
b. Merge $_{\langle \blacklozenge, \clubsuit \rangle}$ (n,m) → n+m+1
c. Merge $_{\langle \blacklozenge, \spadesuit \rangle}$ (n,m) → n+m

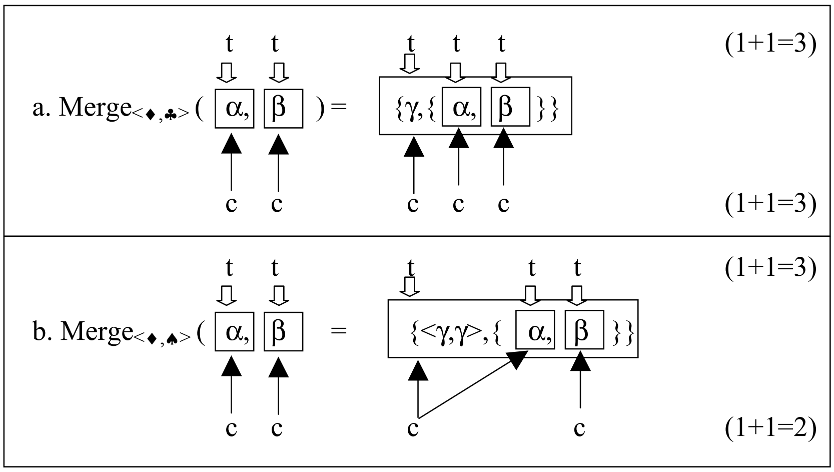
To illustrate this, I prefer to resort to an illustration (t = term, c = category).

---

[159] I say "should" since one mustn't inquire into the structure of the label, that is, the ordered pair $\langle \alpha,\alpha \rangle$. Nunes&Thompson (1998) did inquire further, reducing $\langle \alpha,\alpha \rangle$ to $\{ \{ \alpha \} \}$ on the basis of the Wiener-Kuratowski definition of ordered pairs. A set-theoretic representation of adjunction structures would then look like $\{ \{ \{ \alpha \} \}, \{ \alpha, \beta \} \}$. This requires a sharpening of the definition of terms in order to prevent the unwelcome consequence that singleton α, i.e. $\{ \alpha \}$, becomes a term in adjunction structures over and above α, β, and $\{ \{ \{ \alpha \} \}, \{ \alpha, \beta \} \}$. See Nunes&Thompson (1998) for a solution on the basis of distinguishing dominance and containment. Again, we don't really know which properties of set-theory minimalist syntax is committed to. We have evidence for appeal to the membership relation in the definition of terms (73). Further than that, we have to be cautious given the caveat wrt "full-fledged set-theory" in section 2.6.1. Thus, other formalizations of ordered pairs are possible, as long as the first and second object are uniquely determined (Quine 1959, p.241). One variant of potential interest to minimalist syntax is to define $\langle \alpha,\beta \rangle$ as $\{ \alpha, \{ \alpha, \beta \} \}$ (Cormen et al. 1990, p.80). As far as labels of adjunction structures are concerned, this complicates the picture even further, since $\langle \alpha,\alpha \rangle = \{ \alpha, \{ \alpha \} \}$ gives rise to two additional terms, namely, $\{ \alpha \}$ and α, the latter, of course, identical to the terminal head of the construction. Retroactively, this would make standard Merge operations define ordered pairs (Merge $(\alpha,\beta) = \{ \alpha, \{ \alpha, \beta \} \}$) as long as the projector is the terminal head itself. This was pointed out to me by John Frampton (p.c.). In a way it even makes sense given the asymmetry that the projection relation introduces among the arguments of Merge. One could thus refine the theory further by taking every Merge operation to define an ordered pair. See Saito&Fukui (1998, p.455), who, relying on the asymmetry of ordered pairs, reintroduce linear precedence into the Merge operation in order to keep the head-parameter stateable in syntax. It is, however, not excluded that, in order to steer clear of complications wrt terms, one either takes $\langle \alpha,\alpha \rangle$ as notationally primitive, just distinguishing it from α, or resorts to the theory of "tuples" which takes $\langle\ \rangle$ to be a constant different from $\{\ \}$ (cf. Manna&Waldinger 1985, chapter 12). Avoiding complications involved in labeling, Stabler (1996, 1998) adds two constants, $>$ (left) and $<$ (right) [sic !], to his tree-based version of minimalist syntax. These exhaustively represent the non-terminal vocabulary and function as "pointers," indicating for each constituent on which subtree the terminal head can be found. Thus, $\{ \alpha, \{ \alpha, \beta \} \}$ would be translated as $[_{>} \alpha\ \beta\ ]$, $\{ \beta, \{ \alpha, \beta \} \}$ as $[_{<} \alpha\ \beta\ ]$. Although violating the principle of inclusiveness, Stabler's strategy captures the fact that the featural information operated on is and remains stored in the terminal heads. Cf. section 3.3.3 below.

[160] The difference equally arises for Move, given that Move integrates Merge, in a way not yet fully satisfactorily defined (2.6.2), such that the voyager can create a new specifier by standard Merge or be adjoined to the target.

(89)



This, however, is about as precise as I can, or want to, get. Categories that result from adjunction are non-trivial objects.

It is reasonably clear that one has to define the dominance and c-command relation by means of terms first, in order to subsequently add a linguistically adequate extension of these relations for "segmented" categories, which would allow the usual distinction of "exclusion" (= dominance by no segment), "containment" (= dominance by some segment), and "inclusion" (= dominance by every segment).[161]

Let me point out two further pecularities of adjunction. For one thing, it is an open question to what degree the terms comprising a category formed by adjunction remain independent. Clearly, if Move manipulates terms (cf. (80) above and the subsequent discussion), the lower segment of a category could be separated from the higher one in the course of the derivation. That something like that might be necessary has for example been argued for by Sabel (1996, p.86f, p.226f) on the basis of German incorporation data, given in simplified form below.

(90) a. [daß er die Tür mit dem Schlüssel $[_{PP}\ t_i\ ]\ [_{V°}\ [_{P°}$ auf $]_i\ [_{V°}$ schloß]]]
that he the door with the key open locked
"that he unlocked the door with the key"
b. [Er $[_{V°}$ schloß $]_j$ die Tür mit dem Schlüssel $[_{PP}\ t_i\ ]\ [_{V°}\ [_{P°}$ auf $]_i\ t_j$]]

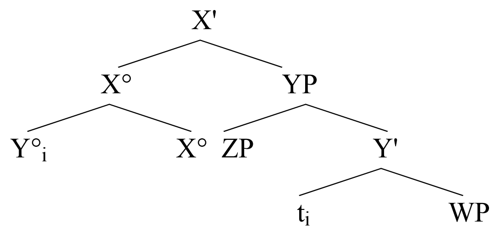
According to that theory, verbal particles have to incorporate into the main verb by adjunction, (90a), unless they topicalize into the specifier of CP. Since, however, finite main verbs undergo verb second movement in root clauses, (90b), these have to break

[161] See Chomsky (1995a, section 4.8) and Nunes&Thompson (1998) for alternative ways of developing the system. See Wartena (1994), Grefe&Kracht (1996), Kolb (1997b), and Kracht (1999), for formalizations of adjunction structures.

up the adjunction structure already formed, movement affecting the lower segment of *V°* only.[162]

Another question concerns the c-command domain of adjoined elements. The simplest version of c-command takes each term to c-command its sister term, i.e. its coargument under Merge, and the terms that sister term dominates. As soon as adjunction structures come into play, it is possible to introduce finer distinctions. Evidence for the right formulations has been sought in the closely related domains of movement theory, binding theory, and the theory of scope. In fact, the most solid need for a special treatment of adjunction comes from movement theory. Assuming that the "voyager" has to c-command its trace, i.e., that movement only goes to c-commanding positions, X°-movement, analyzed as an adjunction operation, poses a problem closely related to the question of cyclicity.

(91)



As can be inferred from (91), t_i neither is the sister term of Y°_i nor is t_i dominated by that sister-term, the lower segment of X°. The standard solution to this problem is to let adjoined elements inherit the c-command domain of the *category* they are adjoined to. Categories are taken to c-command whatever their highest segment c-commands. Thus, Y°_i in (91) inherits the c-command domain of the *category* X°, that domain being YP, and everything dominated by YP. It follows, that Y°_i c-commands t_i as required.[163]

Note, however, that, if (91) has to be the output of Move, it conflicts with hypothesis 3 (of section 2.5.2) in that it requires the addition of a term in a non-root position.

H3: There are no counter-cyclic syntactic operations

Thus, in section 3.4.2 I'll discuss ways of treating of X°-movement in compliance with hypothesis 3.

---

[162] This process is also called "excorporation." The resulting effect of stranding the particle is responsible for descriptively calling them "separable particles." See Sabel (1996) for a wealth of data and intricate discussion of adjunction theory, and Chomsky (1986b) for the original source of much of this discussion. See also den Besten&Webelhuth (1990) for the extractability of VP-segments in the interaction of scrambling to VP and remnant VP fronting.

[163] Another solution can be built on a system of index-sharing as discussed e.g. by Sternefeld (1991, p.137ff). Multiple WH-fronting in certain Slavic languages has sometimes been taken to build adjunction clusters in the specifier of CP, an analysis which can be extended to LF movement of WH-elements (cf. Müller 1995 and references cited there). See Koizumi (1995, p.153ff) for a potential alternative to this controversial analysis, based on multiple specifiers.

More generally, it seems to me that the extended c-command domain for adjoined elements can only be founded on highly theory-internal facts. Thus, most adjoined elements, adverbials in particular, take scope only over what they immediately attach to, in keeping with the simplest version of c-command discussed above. In fact, extending the c-command domain can even produce unwelcome results. Thus, the German focusing particle *nur* ("only"), taken to be restricted to adjoined positions, requires a focused associate in its c-command domain (cf. Bayer 1996). (I annotate focus association by superscripting F).

(92) a. weil [$_{IP}$ Hans [$_{VP}$ nur$^{\mathbf{F}}$ [$_{VP}$ [$_{PP}$ über das Wetter] [$_{V°}$ REDETE]$^{\mathbf{F}}$]]]
because Hans only about the weather talked
"because Hand only TALKED about the weather"
b. ?? weil [$_{IP}$ [$_{DP}$ HANS]$^{\mathbf{F}}$ [$_{VP}$ nur$^{\mathbf{F}}$ [$_{VP}$ [$_{PP}$ über das Wetter] [$_{V°}$ redete]]]]

VP-topicalization, however, does not extend the c-command domain of *nur* to encompass the subject, as (93) shows.[164]

(93) * [$_{CP}$ [$_{VP}$ nur$^{\mathbf{F}}$ [$_{VP}$ über das Wetter geredet]]$_i$ [$_{C'}$ hat [$_{IP}$ [$_{DP}$ HANS]$^{\mathbf{F}}$ t$_i$]]]

Subtle though these facts may be, it can be considered a serious question whether adjunction structures motivate complications of c-command definitions.

Wrapping up the discussion of this section, I will therefore pose two closely related questions.[165]

---

[164] Again, I'm simplifying radically. Note that ungrammaticality of (92b) and (93) concerns just focus association. Thus, neither of the two can be translated as (roughly) (i):
(i) Only Hans is an x, such that x talked about the weather
The strings as such are well-formed if the focused subject is construed outside of the scope of *nur*. (ii) would be compatible with both (92b) and (93).
(ii) Hans is an x, such that only to talk about the weather is a y, such that x y-ed

[165] Doubts about Q4(i) can be further backed up by some rather technical considerations. Extending the c-command domain of adjoined elements is often formulated in a way that dominance by higher segments is simply disregarded. This, however, is clearly not compatible with typical generative treatments of QR, which take (i) and (ii) each to unambiguously represent one of the scopal relations among quantifiers (cf. Huang 1995, p.130).
(i) [$_{IP}$ Everyone$_i$ [$_{IP}$ someone$_k$ [$_{IP}$ t$_i$ loves t$_k$]]] (∀∃) / * (∃∀)
(ii) [$_{IP}$ Someone$_k$ [$_{IP}$ everyone$_i$ [$_{IP}$ t$_i$ loves t$_k$]]] * (∀∃) / (∃∀)
Since the adjoined quantifiers are separated only by a segment of IP, they should c-command each other if c-command domains are extended. See Jacob (1986) for some discussion. Likewise, it is problematic to combine Chomsky's syntactic theory, which allows adjunction to intermediate projections with Kayne's (1994) antisymmetry account of the precedence relation, which requires asymmetric c-command between elements for them to be linearly ordered. Thus, to take the simplest example possible, in (iii), subject and adverbial would symmetrically c-command each other if adjuncts extend their c-command domain in a way disregarding dominance by higher segments of categories.
(iii) [$_{IP}$ They [$_{I'}$ probably [$_{I'}$ didn't realize that]]]

Q4: (i) Do adjoined elements require an extended c-command domain?
(ii) Does adjunction motivate a distinction of categories and terms along with the formulation of two Merge operations?

The preliminary answer given to (i) is negative. Concerning (ii), I suggest that a mechanism along the lines of what Higginbotham (1985) proposes could be enough to single out adjoined and non-adjoined objects. Thus, adjunction could be encoded in checking or saturation relations.[166]

2.6.4 Intermediate Projections and Visibility

In the preceding section, we have seen how endocentricity is added to the minimalist system. However, the operations Select, Merge, and Move do not cover X-bar levels, that is, the distinction between X°, X' (= X^1), and XP (=X^2) projections. Clearly, inclusiveness does not allow the numerical devices that have often been appealed to at least in expositions of X-bar theory, such as (94).

(94) $X^i \rightarrow \ldots X^j \ldots (2 \geq i \geq j \geq 0)$

Still, on a par with the construction of labels, such primitive numerical devices should also be within reach. Thus, syntactic elements could be taken to be ordered pairs of a term and a bar-level indicator, the latter modeled on traditional ways to set-theoretically construct the natural numbers.

(95) a. $< \alpha, \alpha > \ = \alpha°$
b. $< \alpha, \{ \alpha \} > \ = \alpha^1$
c. $< \alpha, \{ \alpha, \{ \alpha \} \} > \ = \alpha^2$

Although such a move would further clutter up notations, it doesn't exceed the formal power of set-theory hitherto employed (cf. section 2.6.3).[167]

Note that cases like (i)-(iii) cause no problem under Epstein's (1999) theory of c-command, which postulates that this relation be fixed once and for all at the point where Merge applies. This allows c-command relations to arise only among co-arguments of Merge operations. Interestingly, X°-movement, as discussed in the text, cannot be captured straightforwardly by that theory.

[166] See Bierwisch (1997) for elaborations on Higginbotham (1985). Cf. Chomsky (1995a, sections 4.7.3 and 4.7.5) and Saito&Fukui (1998) for further discussion of adjunction and minimalist syntax.

[167] It is not even clear whether the objection that this kind of mathematical device is too strong for its task (Zwarts 1992, p.15) really holds of such a construction. What counts as a mathematical device depends on the way it is employed. Thus, as long as no rule refers to multiplication, division, or square roots of bar-levels, the realm of linguistics doesn't seem to have been left. That greater bar-levels dominate smaller ones can be derived from recursive bottom-up application of Merge and its incremental properties. Furthermore, bar-levels might undergo compatibility checks much the same as features from FFORM do under checking theory. Take the most basic perspective of

There have been alternative proposals to deal with bar-levels allowing somewhat greater flexibility. Thus, feature pairs like [+/- projector], [+/- projection] or [+/- minimal], [+/- maximal] could characterize the three basic bar-levels as follows.[168]

(96) a. X° = [+ projector], [- projection] / [+ minimal], [- maximal]
b. X' = [+ projector], [+ projection] / [- minimal], [- maximal]
c. XP = [- projector], [+ projection] / [- minimal], [+maximal]

On the basis of this kind of system, Chomsky (1995a, p.242) provides a contextual definition of bar-level status that observes inclusiveness.

(97) "A category that does not project any further is a maximal projection XP, and one that is not a projection at all is a minimal projection X^{min}, any other is an X', invisible at the interface and for computation."

Under this version, even the missing fourth combination of (96), that is, [- projector], [- projection] / [+ minimal], [+ maximal], is allowed, adding to the flexibility of the system.[169]

Note that crucial functions of X-bar status are formulated wrt to chains and will be taken up in section 2.7.1 below.

The two things that will be discussed here are (i) why bar-level status is defined for categories in (97) and (ii) what "invisibility" means for syntactic computations. Both questions are in fact closely related.

The invisibility of intermediate projections has earlier been stipulated for X-bar systems, primarily on empirical grounds.

(98) "Only minimal and maximal projections (X° and X'') are "visible" for the rule Move-α" (Chomsky 1986b, p.4).

---

microprogramming (Goldschlager&Lister 1988, chapter 4). Checking identity of features could mean to subtract one bit expression from another, successful if zero results, canceled otherwise. Alternatively, bit expressions could be added bitwise, registering only the final bit of each outcome. The features can be taken as checked if nothing but zeroes result, unchecked otherwise. (For example: 0011/0011: 0+0 = 0, 0+0 = 0, 1+1 = 0, 1+1 = 0: √; 0010/0011: 0+0 = 0, 0+0 = 0, 1+1 = 0, 0+1 = 1: *) Nothing prevents a target functor like AgrS°, T°, or AgrO° to require its specifier to be of a specific complexity, e.g. XP, easily translatable into some bit string. If correct, the real question of whether or not bar-levels should be employed is neither inclusiveness nor excessive formal complexity but fruitfulness. Do they serve an essential function or can they be dispensed with? See section 3.3.3, where the above construction of bar-levels is appealed to in order to bring about ("vertical") individuation of nodes.

[168] Cf. Muysken (1982) and Muysken&van Riemsdijk (1986) for an early proposal of such a system and von Stechow&Sternefeld (1988), Kornai&Pullum (1990), Speas (1990), and Zwarts (1992) for further discussion.

[169] Clitics are taken to instantiate that combination (Cf. Chomsky 1995a, p.249).

It follows that intermediate projections can neither be moved nor serve as targets for movement. Together with restrictive phrase-structure rules like the ones in (99), one could derive “second effects“ (100b) and a ban on stranding specifiers (100d).[170]

(99) a. $X^{\alpha} \rightarrow X^{\alpha} Y^{\alpha}$ [$\alpha \in \{ 0, 2 \}$] [adjunction]
b. $X' \rightarrow X° (YP)$ [YP = “complement“]
c. $XP \rightarrow ZP X'$ [ZP = “specifier“]

(100) a. Dieses Buch <u>lese</u> ich morgen
This book read I tomorrow
“This book, I'll read tomorrow“
b. * Dieses Buch morgen <u>lese</u> ich
c. [$_{CP}$ Dieses Buch lese ich morgen]$_i$ glaube ich t_i
This book read I tomorrow think I
“I think I'll read this book tomorrow“
d. * [$_{C'}$ Lese ich morgen]$_i$ glaube ich [$_{CP}$ dieses Buch t_i][171]

Minimalist syntax, however, is developed as a derivational system. Hence, an output of the form in (101) is licensed representationally.

(101) [$_{XP}$ YP [$_{X'}$ ADJ [$_{X'}$ X° ZP]]]

From a derivational perspective, no contradiction arises, given the contextual definition of bar-level status. The adjunction operation is an application of Merge, which takes ADJ and [$_{XP}$ X° ZP] as its input. [X° ZP], being a projection of X° and not having projected any further at that stage, is of the status [- minimal], [+maximal], (= XP). Thus [X° ZP] is “visible“ and the adjunction can occur. Likewise, multiple specifiers enter the realm of possibilities.[172]

It would lead us too far afield if we were to go into an empirical debate on “second-effects“ and possible adjunction sites, relevant to whether the modifications of X-bar theory in question are desirable. “Second-effects“ had been a reason for disallowing them in earlier models (cf. Chomsky 1986b, p.6). The revised line on this is that such effects “may belong to the phonological component“ (Chomsky 1995a, p.368).[173]

Still, invisibility of intermediate projections rules out extractions like (100d). The price to pay, however, is to define bar-level status wrt categories. Again, X°-movement furnishes the most solid reason. Consider the structures in (102).

---

[170] (98) leaves open the possibility of base-generated adjunction to X'. So, additional assumptions are needed to fully derive “second effects“ like (100).

[171] The resulting string of (100d) may actually be acceptable under an analysis where the fronted constituent is a CP the specifier of which has been “topic-dropped.“ The string-final DP would then be a right-dislocated element, providing the topic. Such an analysis puts heavy constraints on context and intonation.

[172] See Koizumi (1995) and the discussion of X°-movement in section 3.4.2 below.

[173] See Zwart (2001) for critical discussion.

(102) a. $[_{Y'}\ Y° [_{ZP} [_{Z'}\ Z°]]] >>> [_{Y'} [_{Y°}\ Z°_i\ Y°] [_{ZP} [_{Z'}\ t_i]]]$
b. $[_{X'}\ X° [_{YP} [_{Y'} [_{Y°}\ Z°_i\ Y°] [_{ZP} [_{Z'}\ t_i]]]]] >>>$
$[_{X'} [_{X°} [_{Y°}\ Z°_i\ Y°]_j\ X°] [_{YP} [_{Y'}\ t_j [_{ZP} [_{Z'}\ t_i]]]]]$

In our case, successive-cyclic X°-movement consists of Z°-to-Y°, (102a), followed by Y°-to-X°, (102b). If bar-level status were computed wrt terms, Move could not operate in the required fashion for (102b). This is so, because the upper segment of Y° is both a projector (of Y') and a projection (of the lower segment of Y°), i.e. it is an intermediate projection and thus "invisible" for C_{HL}. Lifting the definition of bar-levels to categories and assuming that a category is a projector iff its highest term is a projector and a projection iff its lowest segment is a projection immunizes complex X°-clusters against invisibility. The *category* Y° in (102b) is then not a projection, that is, it counts as [+ minimal], and therefore it can undergo movement.[174] This answers our first question.

Now, what about invisibility of intermediate projections? If we take stipulation (98) as indicative of what is meant, a rather unspectacular interpretation suggests itself. Certain objects are not in the domain of certain operations. This requires a proper definition of the operations in question. Thus, single features are not in the domain of Select, which solely applies to members of ℕ. Likewise, neither labels nor single features are operated on by Merge, since it has been defined accordingly. Invisibility as invoked by (97), however, appears to be of a different, more general and non-stipulative, nature. It seems to be taken as a property of the entire system, which doesn't have to be stated in the definition of each operation. What is invisible to the computation cannot be operated on for principled reasons, the question of being in the domain of an operation doesn't even arise.

Again, we have to consider the interfaces for motivation (cf. section 2.2). For intermediate projections the reasoning goes as follows (Chomsky 1995a, p.242).

> "It is also apparent that some larger units constructed of lexical items are accessed, along with their types: noun phrases and verb phrases interpreted, but differently, in terms of their type, and so on. Of the larger units, it seems that only maximal projections are relevant to LF interpretation. Assuming so, bare output conditions make the concepts "minimal and maximal projection" available to C_{HL}. But C_{HL} should be able to access no other projections."[175]

---

[174] Since nothing is said about terms wrt bar-level status and visibility, it looks somewhat unlikely that terms that aren't categories (= lower segments of adjunction structures) can be extracted from their category by movement in the way proposed for (90) of section 2.6.3.

[175] The relation between "bare output conditions" (BOC) and the "Principle of Full Interpretation" (FI), (section 2.2), is not easy to describe for want of sufficient illustration. Having two principles in place, of course, allows some leeway for execution. FI might be stricter in that it leads to ungrammaticality if violated. A lot of subtleties could, however, be hidden here. Thus, although uninterpretable features must be checked in order not to violate FI (cf. section 2.4), it should follow from BOC that they aren't even accessible to C_{HL}, which means they couldn't drive operations,

If tenable, this would elegantly derive stipulation (98) from an interface-oriented perspective. However, some technical and conceptual issues have to be clarified first. It follows from (97) above that being a projection is a property of categories. Thus, invisibility affects categories that are intermediate projections. Now, in section 2.6.3, we concluded that the only objects plausibly called categories in minimalist syntax are the members of $\mathbb{N}$ and the sets constructed from them by Merge and Move. The question therefore arises what it means that a set within a recursive structure of sets is invisible. The safest answer surely is to say that this is not a matter of intuition but definition. Once more, doubts may surface about the commitment of minimalist syntax to even the most rudimentary aspects of set-theory. Yet, set-theory provides the most reliable perspective on the objects defined above, not to speak of occasional explicit appeal to sets and the membership relation (cf section 2.6.2 and 2.6.3). Now, if sets are individuated by their members, that is, if the Axiom of Extensionality holds,[176] it is likely that invisibility of a set has an effect on the visibility of its members. Consider (103).

(103) $\boxed{\{ \alpha, \{ \gamma, \boxed{\{ \alpha, \{ \boxed{\alpha,} \boxed{\beta} \}\}} \}\}}$

If the intermediate category { α, { α, β } } is invisible for C_{HL} this might mean that all of its members and their members are invisible too. Such an unwelcome consequence would be corroborated by our conjecture that the sets operated on in minimalist syntax replace subtrees, not nodes (cf. section 2.6.2 and 2.6.3). Take an application of Move for example.

(104) { . . . { α, { α, β } } . . . . { . . . (COPY { α, { α, β } }) . . . } }

Raising a syntactic element, { α, { α, β } } in our case, leaves behind a copy of the same type. Now, copies are subject to an additional PF-relevant operation

> "[. . .] Delete (Delete-α), which we have assumed to leave the structure unaffected apart from an indication that α is not "visible" at the interface" (Chomsky 1995a, p.250).

Our α in (104) is { α, { α, β } }, (COPY) being an informal notational device.[177] In this case, invisibility requires whatever is inside { α, { α, β } } to be invisible as well, that is, unpronounced at π. It would be detrimental to the theory to allow for a DP like *the*

couldn't be checked, and couldn't be detected by FI, if FI is part of C_{HL}. Certainly an unwelcome result.

[176] For further discussion of this question, see section 3.3.2.

[177] Chomsky (1981, p.89) employed for the same purpose "a feature D indicating that it [= α, H.M.G.] is to be deleted in the PF-component (in fact, D is redundant, determinable from other properties of the grammar)."

toves not to be pronounced at the same time as members of members of that DP, i.e. *the* and *toves* do get pronounced.[178]

If there has to be another type of "invisibility" for intermediate projections, and clearly there has to, otherwise entire subtrees would get inaccessible to C_{HL} in the course of the derivation and no well-formed interface representations could be reached, that alternative type of "invisibility" would have to be properly implemented, providing for among other things definitions of what happens to relations defined wrt to such invisible objects.

In sum, I argue that the kind of invisibility that is supposed to apply to intermediate projections should either be discarded or properly implemented as has, by and large, been the case in the realm of PF-deletion.[179]

On a more conceptual note, one can add that quite a bit of semantic commitment is implied by the assumption that intermediate projections do not get "interpreted" at the interface. In other words, transition from syntactic structures to semantic ones does not have to consider intermediate projections. Such a position is, of course, in conflict with standard compositional semantics, which assigns an interpretation to each subconstituent of a syntactic structure.[180] It conforms more readily to the theory presented in Jackendoff (1990) that "major syntactic phrases correspond to major conceptual constituents" (Jackendoff 1990, p.25). But in this version of "X-bar semantics," which matches "basic formation rules of X-bar syntax", including $X' \rightarrow X$ - *Comp*, with a "basic formation rule for conceptual categories," recursivity of the "argument structure feature" "allows for recursion of conceptual structure and hence an infinite class of possible concepts" (Jackendoff 1990, p.24). Nowhere is it required that "intermediate stages" of the recursion don't count, i.e., that they don't produce "possible concepts."[181]

Moreover, combining the minimalist analysis of possessive structures (105)

---

[178] Nunes (1995) actually takes terms to be deletable while their subterms are left unaffected.

[179] Brody (1998) argues for a version of X-bar theory without intermediate projections, relying on ternary branching instead. Actually, I'm not saying that implementing exactly the right kind of "invisibility" cannot be done. It only will require careful formulation, making more explicit what kind of structure really gets manipulated. A subtle way out might be to employ the layered "ontology" and claim that intermediate projections are invisible qua *category* but visible qua *term*. If so, Move doesn't operate on terms but categories, contrary to the formulation in (80) above. Such a shift, of course, lacks the principled nature of interface induced invisibility and comes close to stipulation (98) again. Secondly, why should terms, and intermediate terms at that, be any more interpretable at the interface than categories. Quite on the contrary, it seems that terms simply provide the *hidden* structure that nodes provide in tree-based formalizations, the latter allegedly suspect for reasons of inclusiveness (2.6.1).

[180] Cf. Partee (1975). Zwarts (1992) worked out a system explicitly designed to combine compositional semantics and X-bar theory, including intermediate projections. See also von Stechow (1993) and Higginbotham (1985, p.554), whose theory recognizes semantic values which are functions of among other things (sub-)phrase-markers, the intermediate projection INFL' among them.

[181] A number of generative theories assign the role of "predication" to intermediate X-bar projections, the specifier holding the "subject" of the predication. See Vogel&Steinbach (1994, chapter 4).

(105) $[_{DP}$ who $[_{D'}$ $[_{D°}$'s] $[_{NP}$ book]]] (Chomsky 1995a, p.263)

with the minimalist treatment of LF reconstruction (Chomsky 1995a, p.202ff, p.290f) one expects structures like (106) to result.

(106) a. Whose brother did they arrest?
b. [Which x, x a person] [they arrested $[_{DP}$ x $[_{D'}$'s brother]]][182]

The bound variable x placed in the specifier of DP is complemented by D', a function of x, invisibility of which, even if technically implemented somehow, shouldn't follow from the fact that D' isn't "interpretable."[183]

Though far from conclusive, I take it as less than established on independent grounds that intermediate projections are uninterpretable. I therefore venture on the following hypothesis.

Hypothesis 4: Invisibility of syntactic elements for C_{HL} has either to be avoided or it must be stipulated and properly implemented (=H4)

H4 further corroborates the "internalist" construal of minimalism as a theory of "narrow syntax." Thus principles like FI, originally designed to constrain syntax from the outside play little or no such role (cf. section 2.2 and 3.1).[184]

---

182 Cf. Chomsky (1977, p.83).

183 See also Engdahl (1986, p.206ff) for an analysis where the ingredients of D' provide one conjunct of a larger quantificational structure. For the interpretability of N', see Huang (1995).

184 If invisibility of intermediate projections is simply abandoned, the status of (multiple) specifiers and the question how to derive linear precedence in terms of Kayne's LCA at PF need some reconsideration.

2.7 Chains and the Power of Copying[185]

Let's assume that syntactic operations and the objects they give rise to can be defined in such a way that the additional challenges discussed in sections 2.5 and 2.6 are met. It is then time to turn to one of the core concerns of generative syntax, namely, the proper treatment of (unbounded) dependencies. In particular, a close analysis of the role of "chains" in minimalist syntax is required. Again, this raises a number of subtle questions having to do with (i) the identity of objects chains are composed of, (ii) the role of chains in defining "legitimate LF objects," and (iii) the X-bar theoretic uniformity condition on chains. I'll deal with these issues in section 2.7.1. Section 2.7.2 will then focus on the tension between use of copies and checking theory. Thus, (i) imposing strictest locality on checking leads to a "resource paradox." On top of this, (ii) intermediate as well as counter-cyclically created positions give rise to what could be called an "explosion problem." Finally, in section 2.7.3 I will give a sketch of how the problems raised can be dissolved on the basis of "multidominance structures." Section 2.7.4 adds an excursus on feature movement, and feature checking without movement.

2.7.1 Identity, Uniformity, and Legitimate LF Objects

The general approach to (unbounded) dependencies hasn't changed much over the years. Devices are employed that define non-local relations between fillers and gaps and thereby encode what, from the perspective of strings, looks like a permutation of (yields of) syntactic constituents. Yet, no consensus has been reached on whether essential constraints on dependencies should refer to successive derivational stages, stressing a procedural approach by means of transformations like Move. Or, alternatively, whether such constraints should be stated wrt a single syntactic representation (like σ or λ of section 2.4), which, given trace theory, encodes the "history" of the derivation.[186]

The minimalist approach appears to be double-edged in this respect. At first sight, commitments are pretty clear.

> "I will continue to assume that the computational system C_{HL} is strictly derivational and that the only output conditions are the bare output conditions determined "from the outside," at the interface" (Chomsky 1995a, p.224).

> "Viewed derivationally, computation typically involves simple steps expressible in terms of natural relations and properties, with the context that makes them natural "wiped out" by later operations, hence not visible in the representations

[185] Part of this section is also presented in Gärtner (1999).

[186] See Chomsky (1995a), Brody (1995), Koster (1987), and Gazdar et al. (1985) for a few influential competitors.

> to which the derivation converges. Thus, in syntax, crucial relations are typically local, but a sequence of operations may yield a representation in which locality is obscured“ (Chomsky 1995a. p.223).[187]

For the purpose of feature-checking (cf. section 2.4), specifier-head and head-head relations are taken to constitute the “natural“ local context. There is no long-distance checking.[188] Thus, a DP in the specifier of VP cannot check nominative case against I° without moving to the specifier of IP. For the purpose of movement, locality means that only the closest element able to check a given feature against a functional target can move there, closeness being definable in terms of c-command and barriers.

Indeed, considerable effort has gone into definitions of locality, conceived of as a condition on the operation Move. Such conditions are taken to be relativized to features and computable at each stage of the derivation at which Move has to apply.[189]

One might expect, then, that a proper definition of locality completes the minimalist account of feature-driven constituent reorderings. Yet, two kinds of factor, one internal, another external to syntax, seem to require a more abstract approach.

First, in order to stick to a universal repertoire of features (and functors) while at the same time preserving word order differences, generative syntacticians assume that certain movement operations are “abstract,“ leading to feature elimination without effects on PF-output. This is encoded in the weak/strong dichotomy of features, ultimately deriving from a universal typology of syntactic positions (cf. section 2.4).

Secondly, there is at least some evidence that certain syntactic elements behave at the interfaces as if they occupied more than one position. Thus, reconstruction configurations at λ are taken to properly distribute the parts of syntactic operators into a tripartite logical representation that distinguishes operator, restrictor, and nucleus, as illustrated in (107), (repeated from section 2.6.4).[190]

(107) a. Whose brother did they arrest?
b. [Which x, x a person] [they arrested $[_{DP}$ x $[_{D'}$'s brother]]]

At PF, some extraction sites have been argued to block the contraction of *want+to* to *wanna* (108c,d), as does an overtly intervening element (108e,f).[191]

---

[187] Proof by example is nontrivial here, given that the copy theory of movement operations fully conserves every step of the derivations, and hence every local context, in the resulting representations. For further discussion, see Epstein et al. (1998) and Cornell (1999).

[188] See, however, Frampton&Gutmann (1999), Chomsky (2000, 2001), and section 2.7.4 below.

[189] See for example Chomsky (1995a), Collins (1997), Lasnik (1995), Müller (1995), and Sabel (1996). Sternefeld (1991) provides a critical discussion of the various theories of locality that still more or less directly influence current versions. For an overview, see also Hornstein&Weinberg (1995).

[190] See section 2.5.2 for binding-theoretic motivation of such a view.

[191] Cf. Chomsky (1981, p.180ff). See Jacobson (1982) for further evidence. I will come back to such PF effects later (see section 3.5).

(108) a. I want to visit Stockholm
b. I wanna visit Stockholm
c. Who_i do you want t_i to visit Stockholm
d.* Who_i do you wanna t_i visit Stockholm
e. I want Mary to visit Stockholm
f.* I wanna Mary visit Stockholm

Thus, "traces" have been taken to occupy the base-positions of moved elements. Abbreviated as *t*, traces were originally analyzed as "empty categories," i.e. phonetically empty terminals $[_\alpha\ e\ ]$ immediately dominated by a node of category α, identical with the category of the moved element (Chomsky 1981, p.90).[192]

As we've already seen in sections 2.5 and 2.6, minimalist syntax construes traces as full *copies* of the "voyager."[193] These can cope more directly with the requirements of reconstruction and binding at λ (Chomsky 1995a, p.202ff). At PF, deletion of phonological content (FPHON) is taken to be a natural option, motivated independently by analyses of ellipsis.[194] Likewise, copying observes the principle of inclusiveness in that nothing but members of LEX or sets constructed from them are made available, no empty terminal *e* having to be postulated.

What makes copying nontrivial, though, are the following two things: (i) the concomitant doubling of resources for the checking machinery, which will be taken up in section 2.7.2, as well as (ii) the questions of identity copies pose. This is where some bookkeeping seems to be unavoidable.

Indeed, to prevent extra complications identical objects could inflict, members of the numeration ℕ are taken to be fully individuated, that is, different tokens of an identical type are "marked as distinct" for C_{HL}.[195] Copies resulting from the operation Move are exceptional wrt individuation, at least so it seems.

> "The operation that raises α introduces α a second time into the syntactic object that is formed by the operation, *the only case in which two terms can be identical*" (Chomsky 1995a, p.251; italics mine, H.M.G.).[196]

Independently, we have to assume that PF-deletion can identify the right object, i.e. the copy left in the base-position (probably aided by c-command relations), so that extra

---

[192] The featural content of empty categories has been the topic of intensive study, sparked off by Chomsky (1981, 1982) and pursued by e.g. Bouchard (1984) and Brody (1984). Earlier versions of trace theory are discussed in Fiengo (1977), and Leuninger (1979, chapter 3).

[193] Chomsky (1981, p.89f) already discussed such an alternative. See Chomsky (2000, p.145fn.62) for a remark on the "historical" status of copy theory.

[194] See Wilder (1995, 1996) for interesting applications.

[195] Cf. Chomsky (1995a, p.227). See section 2.6.1 for discussion of how rich the lexical types might be and what consequences various options have for the structure of ℕ. Note also that something like indices appears to be appealed to in this case.

[196] Recall from section 2.6.3 that labels and terminal heads were taken to be "identical" too, although in a sense that didn't seem to merit further comment.

markings like *(COPY)* remain purely expository devices not interfering with identity (cf. section 2.6.4). But saying that, i.e. talking by means of definite descriptions of “the right object“ or “the copy,“ one has to realize that identity is not a trivial addition to C_{HL}. Take the definition of “standard identity“ (Hodges 1983, p.68).

(109) *a* = *b* iff *a* and *b* are one and the same thing[197]

If what is involved in Move is “one and the same thing,“ talk of “raising,“ “copying,“ “trace“ and “two terms“ is pretty confusing, not the least because of the quasi-physical connotations involved in most of these notions. Clearly, for the output of Move, α counts as two terms, as the term record (77) deriving from definition (73) indicates. Take another simple example. (Boxes correspond to terms.)

(110) a. $\boxed{\{\ \alpha, \{\boxed{\alpha,}\ \boxed{\{\ \beta,\ \{\boxed{\beta,}\ \boxed{\gamma}\ \}\ \}}\ \}\ \}}$

b. $\boxed{\{\ \alpha,\{\ \boxed{\gamma,}\ \boxed{\{\ \alpha,\ \{\boxed{\alpha,}\ \boxed{\{\ \beta,\ \{\boxed{\beta,}\ \boxed{\gamma}\}\ \}}\ \}\ \}}\ \}\ \}}$

(110b) is the result of applying Move in (110a), raising γ. 5 terms of the input map into 7 terms of the output. Clearly, γ counts twice as long as the membership relation is not defined in any different way.[198] Thus, it is not qua *terms* that one can speak of the voyager and its trace as “one and the same object.“ Nor can this be stated by means of the membership relation which is the basis for the computation of terms.[199] Therefore, identity in the sense of (109) would have to be defined over a domain of objects not yet recognized in the theory.[200]

---

[197] An even sharper formulation of the argument developed in the following can be built on “Leibniz's principle.“
(i) If x and y are identical, then y has every property of x (Smullyan, 1971, p.35).
Thus, properties like “being the specifier of IP“ vs. “being the specifier of VP“ distinguish elements of a subject chain in English, which therefore could not count as identical under Leibniz's principle.

[198] This result would have to be confronted with the issue of “extensionality“ raised in section 3.3.2.

[199] Even appeal to categories wouldn't produce a different result, since the category record equals the term record for at least standard Move, i.e. as long as adjunction isn't considered. See section 2.6.3.

[200] A graph-theoretic interpretation of set-theory will be discussed in section 3.3. Such an interpretation, however, is by no means trivial and requires some careful embedding into linguistic theory.

Indeed, there is further evidence that identity in the strict sense can't really be upheld. First of all, Move, as already indicated in 2.6, is taken to introduce another (higher-order) kind of object.

> “The operation Move forms the chain CH = α,t(α), t(α) the trace of α“ (Chomsky 1995a, p.250).

> “Under the copy theory of movement [. . .], a two-element chain is a pair <α,β> where α = β“ (Chomsky 1995a, p.251).[201]

Now, chains are familiar from GB theory, which attributed properties like Θ-roles and Case to chains, that is, to the abstract objects that represent the “derivational history“ (Chomsky 1981, p.45) of an element at S-structure or LF. What makes the notion of chain difficult to integrate into the purportedly derivational minimalist syntax is that it is usually defined over representations in terms of non-local concepts such as *binding*. Take the following characterization.

> “Intuitively, a *chain* is a sequence of categories at S-structure coindexed by Move-α, each member except the first being a trace of the first member, which we will call the *head* of the chain“ (Chomsky 1981, p.331).

Tentatively, one could adapt the LGB definition of chains (Chomsky 1981, p.333) in the following way.

(111) The sequence CH = ($\alpha_1, \ldots, \alpha_n$) of terms (categories) α is a *chain* iff
for each i ($1 \leq i < n$), α_i locally binds α_{i+1}

(112) A term (category) α locally binds a term (category) β iff
(i) α c-commands β, and
(ii) α and β are not marked as distinct, and
(iii) there is no term (category) γ, such that γ, α, and β are not marked as distinct and α c-commands γ, and γ c-commands β

Assuming, as has been done above, that distinctness is always ensured except for the case of copying, Move could be said to “define“ chains in the same sense in that it contributed coindexation in GB, because Move leads to a multiplication of non-distinct terms. Although the exact interpretation of non-distinctness would have to be clarified,

[201] Intended to be a nontrivial use of identity, =, this merits closer attention. Indeed, we need a two-layered definition of “occurrences.“ Thus, α and β in the quote above could be taken as two occurrences of one member from ℕ. On this level, occurrences are not especially marked as distinct. Inside ℕ, however, we can additionally get occurrences of one member from LEX. These latter ones are taken to be marked as distinct, presumably by indexation. See Kitahara (2000) for discussion of how to avoid indexation in the latter case.

a definition like (111) presupposes the *non-identity* of terms in a chain, at least for non-trivial chains. This is corroborated by the following remark.[202]

> "[. . .] we want to distinguish the two elements of the chain CH formed by this operation" (Chomsky 1995a, p.251).

This is done by identifying each element of a chain with its co-constituent.

> "We can take the chain CH that is the object interpreted at LF to be the pair of positions. [. . .] POS_1 and POS_2 are distinct objects constituting the chain CH = < POS_1 , POS_2 > formed by the operation" (Chomsky 1995a, p.252).

Applied to our example (110b) we would get (113), the pair of sister terms of γ and its copy.

(113) $CH = < \{ \alpha, \{ \alpha, \{ \beta, \{ \beta, \gamma \} \} \} \}, \beta >$[203]

Now, something like definition (111) could surely work for characterizations of the interface representation λ. This would come close to Rizzi's (1986) chain formation algorithm. No claim, however, should be made that the concept of chains is particularly simple, minimal, or well-adapted to a purely derivationalist perspective on C_{HL}. Thus, chains certainly do not belong to the domain of simple steps in the sense of section 1.1. Without coding tricks, each operation involving chains during a derivation will have to repeat the non-local computations of (111)/(112). Alternatively, a transparent solution would be to recognize "chain structure" (Rizzi 1986, p.67) and take syntactic structures to actually be pairs $<S_{CON}, S_{CH}>$ of "constituent structure" and "chain structure" on which simultaneous operations are defined (cf. Kracht 2001). This would formally come close to the <C,F> pairs, constituent structure and "functional structure," respectively, postulated in "lexical-functional grammar" (LFG).[204]

---

[202] See section 3.3.2 for further discussion, and 3.4.1 for additional indirect evidence that non-identity is indeed the intended interpretation of copies in minimalist syntax.

[203] Formalization of chains had been sidestepped in Chomsky (1981, p.47f), although some indication was given as to how it could be achieved. See Kracht (2001) for such a formalization. Within GB tradition, discussion of chains generally refers back to Chomsky (1981). The technique of individuating syntactic elements by means of their context goes back to the definition of occurrences in strings in terms of a three-place relation. "Z is an occurrence of X in Y, if there is a W_1, W_2 such that $Y = W_1 + X + W_2 = Z + W_2$" (Chomsky 1975/1955[1], p.110). For example, taking *Y = New + York + City + is + in + New + York + State* and *X = New + York*, *Z = New + York + City + is + in + New + York* is an occurrence of X in Y. See Chomsky (2000, p.114ff) for additional discussion of occurrences.

[204] See Abeillé (1993) and Horrocks (1987) for introductions to LFG, which was developed by Joan Bresnan and Ronald Kaplan. The following remarks in discussing configurational properties of syntactic elements might indicate that such a dissociation of structures is not too far off the mark: "[. . .] the chain CH is not in a configuration at all [. . .]." "But CH is not in any configuration [. . .]" (Chomsky 1995a, p.313).

In order to avoid such extra complexities, one might want to conclude that chains, though virtually present in the form of copies, are an interface phenomenon not relevant to the syntactic computations proper. Recall, however, the minimalist strategy of tailoring syntax to the needs of the interface (cf. 2.2, 2.6.1, and 2.6.4). Chains play a major role in characterizing the notion of "interpretability," crucially underlying the principle of Full Interpretation (FI).

> "At LF, we assume each legitimate object to be a chain CH = (α_1, . . ., α_n) [. . .]" (Chomsky 1995a, p.194).

The only additional condition on chains is that they be uniform or of an operator-variable type (Chomsky 1995a, p.154, p.91; slightly adapted).

(114) Legitimate LF Elements (each a chain)
a. Arguments: each element is in an A-position, α_1 Case-marked and α_n Θ-marked
b. Adjuncts: each element is in an A'-position
c. Lexical elements: each element is in an X° position
d. Predicates, possibly predicate chains if there is predicate raising
e. Operator-variable constructions (α,β), where α is in an A'-position and β heads a legitimate uniform chain

Note that (114a-d) are compatible with sequences of length 1, i.e. so called "trivial chains." Thus, to the extent that minimalist syntax is interface-oriented, chain formation is a necessity, given this typology of legitimate LF elements. But what would be the difference between a chain formation algorithm a la Rizzi, defined on λ, and chain formation during computations toward λ?

The answer to this question brings out further subtleties. Recall from section 2.6.4 that the invisibility of intermediate projections for C_{HL}, problematic though it turned out to be, was motivated by the assumption that such projections weren't interpretable at λ. A strict transposition of this kind of argument would lead one to expect that only chains, i.e. the only legitimate LF elements from (114), are interpreted at λ. It would follow that what isn't a chain isn't accessible to C_{HL} either. Consequently no derivations whatsoever could get off the ground. This is so because chains formally are of a type different from terms (categories), namely, they are *sequences* of terms (categories) from all we have been able to establish so far. Form Chain as a subpart of Move would come too late, given that invisibility of non-chains automatically voids that operation. Clearly, such a result is unwelcome and unintended. If anything, however, this strengthens hypothesis H4, i.e., one should be wary of loosely defined conceptions like invisibility.[205]

[205] That this is not an unfair overstatement of the case can be inferred from the fact that the typology in (114) has been used to motivate the elimination of intermediate A'-traces of operator-variable chains resulting from argument movement. Hence, (114) was taken to derive Lasnik&Saito's (1992) earlier stipulation to the same effect (Chomsky [with Howard Lasnik] 1995a, p.88-92). The

On the other hand, if chains are the only legitimate LF objects, these must come into being somehow. Now, should an explicit derivational operation for chain formation be required, it could not be just Move, which is adequate for the formation of non-trivial chains only. In this case, Rizzi's algorithm, that "reads off" chains from representations (Rizzi 1986, p.66) appears to be more straightforward. Otherwise, there has to be an extra operation called "Form Chain" over and above the one integrated into Move, which brings about some kind of type-shifting from *terms* to a trivial *sequence* of terms, a slightly suspect assumption.[206]

Interestingly, syntactic elements are occasionally identified with chains right away. Thus, the definition of domains is said to be

> "defined "once and for all" for each CH: at the point of lexical insertion for *CH* = α, and when CH is formed by movement for nontrivial CH" (Chomsky 1995a, p.299; italics mine, H.M.G.).[207]

---

required deletion operation is licensed because it leads to the formation of a legitimate LF-object, namely, a pair (α,β) of (114e). "[. . .] operations in general are permissible only to form a legitimate LF object." (Chomsky [with Howard Lasnik] 1995a, p.91, cf. Chomsky 1995a, p. 301). On strict application of this principle, even Merge, which produces *terms (categories)* not *chains*, would not be a "permissible operation," it seems. Note, incidentally, that deletion of intermediate traces isn't fully welcome in the above-mentioned case, at least if the treatment of LF-anaphor raising is to be retained. Thus, the fact that *herself* in (i) can be bound from either the matrix or the embedded subject position, has been taken to show that the anaphor can LF-raise to one of the two I° nodes by X°-movement from either the lowest or the intermediate WH-copy, respectively (cf. Chomsky 1995a, p.208f.).

(i) $[_{CP}$ [Which picture of herself$_{i/k}$] did Mary$_i$ think $[_{CP}$ $(_{COPY}$ [which picture of herself$_{i/k}$]) that
Gloria$_k$ liked $(_{COPY}$ [which picture of herself$_{i/k}$])]]

If, however, the intermediate copy has to be deleted, raising from there to the matrix I° should either be prevented or at least it shouldn't be interpretable since the anaphor wouldn't bind a trace, its extraction site having been eliminated. The problem was noted and assumed to be avoidable under "a slight variation of this approach" (Chomsky 1995a, p.388fn.75). Thus, essentially no deletion is assumed to take place in the end. Consequently, an intermediate trace of an argument, which is not considered part of a legitimate LF object by (114e), is neither invisible at LF, allowing extraction of the anaphor, nor uninterpretable at λ, since it appears to be hosting the restriction on the semantic variable left by anaphor extraction. See also Chomsky (2000, p.146fn.66). Let me add that I take example (ii) to show that the entire approach may need revision.

(ii) ?? $[_{CP}$ [Which picture of herself$_i$] did Mary$_i$ wonder $[_{CP}$ why John liked
$(_{COPY}$ [which picture of herself$_i$])]]

It is plausible to assume that the degraded nature of (ii) is due to an island violation, not to a binding problem. However, since WH-extraction out of a WH-island does not create an intermediate copy, the binding relation between *Mary* and *herself* cannot be brought about by LF-anaphor raising.

[206] Collins (1997, p.90) assumes that "[. . .] "lexical insertion" is a kind of movement from the lexicon." Although this might be interpreted in a way that chains are there in syntax right from the beginning, the problem of accessibility is only pushed one step further down into the lexicon, at least so it seems.

[207] Cf. Chomsky (1995a, p.179).

Here it looks as if terms (categories) are chains right from the start, that is, on being inserted into the derivation. Technically, this means that such objects are *terms* and *sequences of terms* of length 1 at the same time.[208] We are therefore driven to acknowledge a third layer of types/entities on top of "syntactic objects" and "terms," namely, "chains."

However, the question of visibility still isn't fully settled. Being a chain is only a precondition to becoming a legitimate LF object. Thus, if only legitimate (interpretable) LF objects are accessible to C_{HL}, trivial argument chains, whose head is not in a Case position by the logic of checking theory, should still not be visible, as can be verified from (114a). Thus, neither (deep) subjects nor (deep) objects could ever be operated on once inserted into the derivation. Of course, a less strict version of visibility won't permit such counterintuitive conclusions.[209] As in section 2.6.4 above, we arrive at a skeptical perspective on what direct, non-trivial, influence interface conditions could have on the inner workings of C_{HL}.[210]

Now, chains have been appealed to in resolving another technical issue. We have seen in section 2.6.4 that minimalist syntax defines X-bar levels contextually, bar-levels as entities being considered suspect from the perspective of inclusiveness. Thus, the following principle is put into place (=(97)).

(115) "A category that does not project any further is a maximal projection XP, and one that is not a projection at all is a minimal projection X^{min}, any other is an X', invisible at the interface and for computation" (Chomsky 1995a, p.242).

Given that movement alters the context of the voyager by definition, the question arises whether the structure-preserving nature of movement can be ensured or has to be given up. Earlier, the desired results were obtained by stipulation.

(116) a. Only X° can move to the head position
b. Only a maximal projection can move to the specifier position
(Chomsky 1986b, p.4)[211]

In minimalist syntax, such principles should ideally follow on independent grounds. For the sake of brevity, we'll assume that intermediate projections are ignored (cf. section 2.6.4) and that it is invariably the target that projects when Move is applied. We are then left with the following paradigm to handle. We want maximal projections to move

[208] Earlier formulations might be taken to imply a more abstract relation between chains and the syntactic elements they involve. Thus, a chain was said to be "associated with" for example an NP (Chomsky 1981, p.45). The assumption that "chains are not introduced by selection from the lexicon or by Merge" (Chomsky 1995a, p.316) clearly justifies the qualms raised above.

[209] See for all of this Chomsky (1995a, p.299ff.), where legitimacy and visibility are indeed taken to be serious questions for chain formation, the "Chain Condition" (≈ (114a)) being explicitly invoked.

[210] This skepticism seems to be shared by Chomsky (2001, p.49fn.69)

[211] Structure preservation is held to be true of adjunction as well. Thus, X° adjoins to Y°, XP to YP (cf. Chomsky 1986b, p.73), and section 2.6.4 above.

into specifiers (117a) and X°-elements to adjoin to other X°-elements (117d). Conversely, we want to rule out maximal projections adjoining to X°-elements (117b), and X°-elements moved into specifier position (117c).[212]

(117) a. $[_{YP}$ [- min / + max $]_i$ $[_{Y'}$ Y° $[_{ZP}$ . . . ($_{COPY}$ [- min / + max $]_i$) . . .]]]
b. * $[_{YP}$ $[_{Y°}$ [- min / + max $]_i$ Y°] $[_{ZP}$ . . . ($_{COPY}$ [- min / + max $]_i$) . . .]]
c. * $[_{YP}$ [+ min / + max $]_i$ $[_{Y'}$ Y° $[_{ZP}$ . . . ($_{COPY}$ [+ min / - max $]_i$) . . .]]]
d. $[_{YP}$ $[_{Y°}$ [+ min / + max $]_i$ Y°] $[_{ZP}$ . . . ($_{COPY}$ [+ min / - max $]_i$) . . .]]

Concentrating on how to rule out configuration (117c), Chomsky (1995a, p.253) proposes the following condition.

(118) A chain is uniform with regard to phrase structure status.

Indeed, a projector undergoing movement will necessarily alter phrase structure status in the target position, given that it is the target that projects. Thus, [- max] will become [+ max]. On the basis of (118), X°-movement to specifier position cannot be well-formed. Likewise, (117a) conforms with (118), since movement conserves any [- projector] status, that is, [+ max] remains [+ max].

Clearly, however, uniformity of chains as it stands is both too weak and too strong for paradigm (117). In fact, it captures exactly half of those cases. Hence, in order to rule out (117b), another condition is appealed to.

(119) Morphology deals only with X° categories and their features
(Chomsky 1995a, p.319)

(119) is supplemented by a comment.

> "On this natural assumption, the largest phrases entering Morphology are X°s, and if some larger unit appears within X°, the derivation crashes" (Chomsky 1995a, p.319).

Taken as a ban on [- min] items within X° categories, this would properly derive the contrast between (117b) and (117d).[213] However, morphology operates on the PF

---

[212] Borderline cases of syntactic elements that are [+ min/ + max] even in their base-position constitute a complication that need not concern us here.

[213] Here it is important that (115) is defined on categories, not terms. The two-segment category Y° in (117b)/(117d) will count as [+ min / + max] and therefore rightly enter morphology. Though probably not serious, there are matters of fact to be addressed if (119) is to be assumed. Thus, certain nominal constructions appear to require a CP to occur inside N°.

(i) diese $[_{N°}$ $[_{CP}$ jeder-sorgt-für-sich-selbst] Einstellung]
this everyone-provides-for-oneself attitude

One argument for taking CP to be inside N° is the fact that the resulting structure has exactly the distribution of N°. Thus, adjectival modifiers precede CP but cannot intervene between CP and N.

branch of C_{HL}. At λ, where the uniformity of chains most plausibly is going to be checked, (117d) would violate condition (118), incorrectly predicting ungrammaticality.[214] Thus, another filtering device is required.

(120) At LF, X° is submitted to independent word interpretation processes WI where WI ignores principles of C_{HL}, within X° (Chomsky 1995a, p.322).

I will not go into details of implementation, noting only that even together with the morphological filters, reference to chains in (118) can be construed as a condition on λ. Consequently, a need to operate on chains during computation toward π and λ doesn't arise in the area of determining X-bar status and its impact on structure preservation, which means that the concept of chains is probably not indispensable for this account.

2.7.2 Locality, Countercyclicity, and Checking Resources

Although neither Merge nor Move operates on chains, appeal to chains during a computation may be unavoidable for very prosaic reasons. Consider checking theory. It is one of the core assumptions of minimalist syntax that checking is local, i.e. it affects the terminal head X° of the target and whatever element is inserted into its "checking domain," i.e. specifier of XP or adjoined to X°.[215] Under movement, of course, only one term is introduced into the checking domain, the copy staying in the base position. Thus, at first sight, feature-checking by Move is unable to effectively get rid of the features to be checked, each application of the operation doubling half of the resources

(ii) diese egoistische jeder-sorgt-für-sich-selbst Einstellung
(iii) * diese jeder-sorgt-für-sich-selbst egoistische Einstellung
Secondly, CP induces the same stress pattern that compounding results in.
(iv) diese jeder-sorgt-für-sich-SELBST Einstellung (v) diese EgoISTeneinstellung
The main stress, marked by capitals, falls into CP in (iv) as well as into the incorporated nominal element in (v). Standard syntactic modifiers, like adjectives and prenominal genitives, however, are less than or equally prominent as the head N°, unless a contrastive reading is intended.
(vi) diese egoistische EINstellung (vii) Peters EINstellung
See, however, Wiese (1996), who argues for an extra mechanism such as quotation to deal with the above kind of fact. Still, it is open how word-internal bar-level status would be computed. Contextual definitions may not be the ideal mechanism to achieve this. Thus, if the syntactic filter above is based on the "absence in morphology of the syntactic notion "maximal projection"," then "[a] precise meaning must be given to the term "maximal projection." Morphology does have maximal projections in a trivial sense: in the compound *off-white paint*, the adjective *off-white* is a maximal projection of the adjective *white* in the sense that there is no larger projection of the adjective in this compound" (Williams 1989, p.282). What makes a difference, according to Williams, is not context but "intrinsic" notions like case-marking, predication, reference, and opacity.

214 (117b) need not concern us any further if indeed only features undergo movement after Spell-Out.

215 See, however, Frampton&Gutmann (1999), and Chomsky (2000, 2001), as well as section 2.7.4 below.

to be checked. Omitting details, we have a target α, possessing an unchecked feature $[_{\alpha}$ *F] while the voyager β is equally equipped with $[_{\beta}$ F*], *F* standing for whatever structure is necessary for something to constitute a feature, while * signifies the property of being unchecked. Before the two features can cancel, producing $[_{\alpha}$ F] and $[_{\beta}$ F], a copy of β is created, which again possesses F*, i.e. ($_{\text{COPY}}$ $[_{\beta}$ F*]). In fact, after local checking has occurred for the two "originals," nothing is left for the unchecked feature of the copy to cancel against, much to the detriment of the entire approach. Call this the "resource paradox." (121) presents it in schematic detail.

(121) a. [. . . *F . . . [. . . [F*] . . .]]
b. [[F*] *F . . . [. . . ($_{\text{COPY}}$[F*]) . . .]]
c. [[F] F . . . [. . . ($_{\text{COPY}}$[F*]) . . .]]

After checking, the system goes into state (121c). Without extra assumptions, then, the copied unchecked feature, F*, should prevent convergence at one or both interfaces. This problem has been (foot-)noted.[216]

> "Technical questions arise about the identity of α and its trace t(α) after a feature of α has been checked. The simplest assumption is that the features of a chain are considered a unit: if one is affected by an operation, all are" (Chomsky 1995a, p.381fn.12).

The "simplest assumption" presupposes, and therefore provides further evidence for, the non-identity of α and its trace, i.e. they are *not* "one and the same thing" (cf. (109)), although its consequences would be more perspicuous if there was strict identity, i.e. if no copy or trace had been distinguished in the first place.[217]

Yet, keeping the concept of chains as it was introduced above seems to require non-local operations in the sense that features have to be eliminated in the "head" as well as in the "foot" of the chain. Opening this possibility weakens the rigid locality of checking theory quite considerably, although enforcement of locality used to provide the most solid motivation for a Move operation, chains being second-born creatures in that respect. Moreover, identifying (and simultaneously checking) the features across an

[216] Cf. Collins (1997, p.3).

[217] Stabler (1996) designs a version of minimalist syntax that doesn't allow copying of checking resources. My own proposal, to be given in section 3.3, will likewise be designed to avoid copying the members of LEX, such that checking of features always reduces resources directly. This will essentially be achieved by keeping these elements at a fixed "address" defining additional structure over non-terminal elements that only "point" to the members of LEX. Nunes (1995) turns an apparent vice into a virtue and retains unchecked features in copies for deletion at PF. From this he derives the fact that movement before Spell-Out results in pronunciation of the head of a chain. The account, however, relies on an intricate system of different deletion rules, both at PF and LF, plus, above all, a principle of transderivational economy, which, as is going to be argued in section 2.8, I object to. See Nunes (1999) for a shorter exposition and Gärtner (1998) for more detailed analysis of that system.

entire chain clearly requires a richer concept of syntactic structures than hitherto put into place.[218]

Now, the unwelcome “power of copying“ just discussed returns in a different guise with intermediate positions (“intermediate traces“) and countercyclicity. Starting off with the latter issue, let's turn to weak movement targets, the syntax-internal motivation for a more abstract approach to constituent reorderings. Recall from section 2.4 and 2.5.2 that movement of direct objects to the specifier of AgrOP in SVO languages like English and Swedish was supposed to occur after Spell-Out has triggered PF-computations.[219] Thus, this kind of object-shift won't have effects on the pronounceable string at π. Completion of the computation toward λ will then require a counter-cyclic adjustment, given that at least TP and AgrSP must already have been built on top of AgrOP. The step converting (122a) into (122b) (repeated from 2.5.2) occurs after Spell-Out.

(122) a. $[_{AgrSP}$ Kermit$_i$ AgrS° $[_{AgrOP}$ ∅ AgrO° $[_{VP}$ t$_i$ likes beans]]]
b. $[_{AgrSP}$ Kermit$_i$ AgrS° $[_{AgrOP}$ beans$_j$ AgrO° $[_{VP}$ t$_i$ likes t$_j$]]]

We already saw in section 2.4 that what prevents objects from moving to the specifier of AgrOP before Spell-Out in the languages that possess weak features for that position is

> “[. . .] a natural economy condition: LF movement is “cheaper“ than overt movement (call the principle *Procrastinate*). [. . .] The intuitive idea is that LF operations are a kind of “wired in“ reflex, operating mechanically beyond any directly observable effects. They are less costly than overt operations. The system tries to reach PF “as fast as possible,“ minimizing overt syntax“ (Chomsky 1995a, p.198).

However, in the case of WH-movement of objects this potentially leads to a decision problem. Consider the paradigm in (123).

(123) a. $[_{AgrO'}$ AgrO° $[_{VP}$ Johan älskar vad]]
John likes what
b. $[_{AgrOP}$ vad $[_{AgrO'}$ AgrO° $[_{VP}$ Johan älskar $(_{COPY}$ vad)]]]
c. Vad älskar Johan
d. Vad sade Cecilia att Johan älskar
What said Cecilia that John likes
e. Vem sade att Johan älskar vad
Who said that John likes what
f. * Vem sade att Johan vad älskar

---

[218] See Kracht (2001) for a formal analysis of chains.
[219] See Lasnik (1999a) for a different view.

At stage (123a), it has to be decided whether to proceed to (123b) or not. This, however, depends on the global context, i.e., on what other elements are available in $\mathbb{N}$. Thus, if (123c) or (123d) are aimed for, step (123b) is licensed, because *vad* is going to raise to the matrix operator position later. If, on the other hand (123e) has to be constructed, in which there is a matrix operator *vem*, step (123b) mustn't take place, as the ungrammaticality of (123f), taken to represent PF order, shows. This is so, because raising of *vad* to the matrix operator position cannot occur before Spell-Out. Computation of these results, however, is non-local, having to complete alternative derivations in order to determine which ones converge and which of the convergent ones require fewer operations before Spell-Out. Hence, in the case of (123e) and (123f), both strings at π derive from convergent derivations, while (123e) requires one pre-Spell-Out operation less. This could be taken to effectively block the derivation leading to (123f).

Since I'm not convinced that economy principles should be employed in the model of syntax developed here, as will be argued in section 2.8, I'd prefer a version of syntax that deals with covert movement differently.[220]

Yet, I said (122) leads to a *potential* decision problem. Chomsky (1995a, p.302f) seems to countenance the possibility that it is the trace of the operator that does the checking movement after Spell-Out. Consider (124).[221]

(124) a. [$_{CP}$ *What* did [$_{AgrSP}$ John [$_{AgrOP}$ AgrO° [$_{VP}$ ($_{COPY}$ John) see ($_{COPY}$ *what*)]]]]
b. [$_{CP}$ What did [$_{AgrSP}$ John [$_{AgrOP}$ ($_{COPY}$ *what*) AgrO° [$_{VP}$ ($_{COPY}$ John) see ($_{COPY}$ ($_{COPY}$ *what*))]]]]

The notational device indicating copies needn't worry us, as already mentioned above. Recall, however, that the object in the specifier of CP will contain an unchecked Case feature as well. Thus, over and above a solution for the standard "resource paradox," arising from the copies of *what* in VP and Spec,AgrOP, an additional mechanism, presumably based on "chain composition," would have to take care of the unchecked features of *what* in Spec,CP. Clearly, disallowing the skipping of checking positions and concomitant counter-cyclic readjustments would void this extra problem. Again, this accords with hypothesis 3, arrived at in section 2.5.2.

In fact, there is further evidence pointing in the same direction. Thus, consider (125).

(125) a. [[That Kermit likes beans] seems [t'''' to be likely [t''' to appear [t'' to be believed [t' to t sound ridiculous]]]]]]
b. t = ($_{COPY}$ [That Kermit likes beans])

[220] Stabler (1996) likewise appeals to Procrastinate for the treatment of WH-objects in English. His account is backed up by the claim that "economy conditions may have natural implementations in the performance model." To the extent that this is accurate, we have here a strong vindication of "Frampton's conjecture" of section 1.1 above.

[221] The discussion is inconclusive actually. Chomsky (1995a, p.304) introduces the following principle, which would exclude this kind of derivation.
(i) Trace is immobile.

Any such overt movement of a complex constituent would leave a number of full copies, indicated as traces *t* here, the content of which is given in (125b). In the case at hand, each copy contains the unchecked features of the direct object *beans*. This time, however, not even considering the features of the CP-chain as a unit, as in the (problematic) approach to the "resource paradox" would seem to be sufficient. What is required is chain-formation inside of each copy of an already existing chain. Again, this clearly presupposes the recognition of structures richer than the tree-like objects C_{HL} is often taken to manipulate. Chain structures as discussed in Rizzi (1986) and formalized in Kracht (2001) would be a candidate. On such chain structures one would have to define a mechanism for keeping track of and operating on the appropriate counterparts across copies.

The "local" alternative of getting rid of unchecked features inside of each copy individually runs into what could be called an "explosion problem," given that due to recursivity objects copied may be unboundedly complex. Clearly, in order to avoid such an "explosion problem" together with the "resource paradox," some way of restraining the power of copying must be found. Although, this will be the topic of section 3, I'll give a brief sketch of my proposal in the following section.

2.7.3 Restraining the Power of Copying

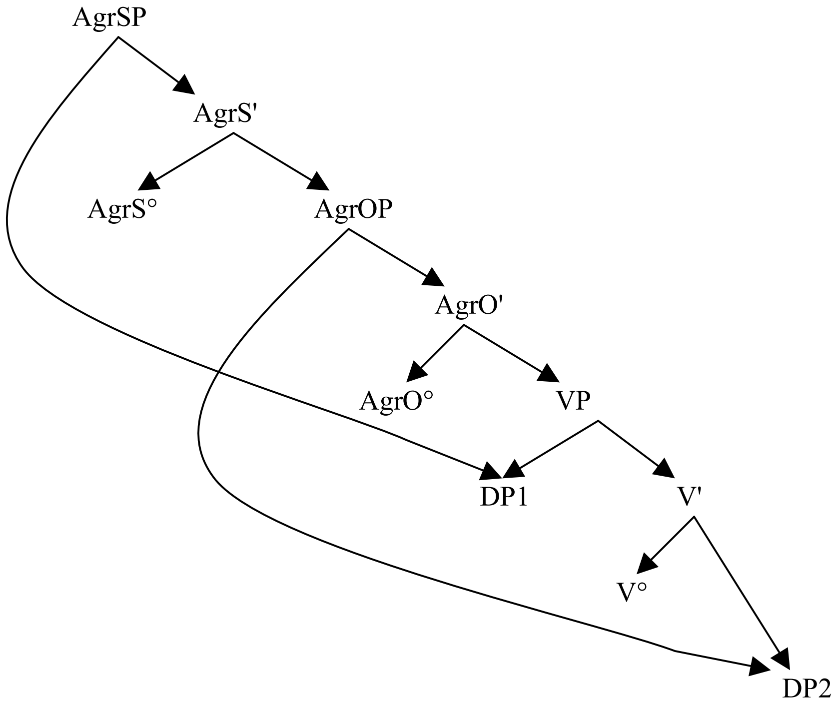
The most radical way of restraining the power of copying would be to simply deny copies any role in minimalist syntax. Thus, syntactic structures can be taken to allow "multiconstituency" or "multidominance." Formally, this involves a shift from tree-like objects to graphs that violate the "Single Mother Condition" (SMC), given in (126). (ID = "immediately dominates"; N is the set of nodes).[222]

(126) Single Mother Condition (SMC)
$(\forall x, y, z \in N)[\, (xIDz \wedge yIDz) \rightarrow x = y\,]$

Although technical discussion will be put off until secton 3.2, the structure in (127) should suffice to convey an intuition of the possibilities offered by such a shift.

[222] See Blevins (1990, p.48) and Barker&Pullum (1990, p.20). McCawley (1968, p.244) called this the "Nonlooping Condition." I'm greatly indebted to John Frampton (p.c.) for bringing this kind of approach to unbounded dependencies to my attention. Precursors and variants will be dealt with in section 3.2.

(127)



Thus, instead of putting a copy of e.g. the direct object in Spec,AgrOP, that constituent will be immediately dominated by both V' and AgrOP. The operation Move will then consist of defining the required “links.“ Given that only a single “copy“ of each constitutent will be recognized, neither a “resource paradox“ nor an “explosion problem“ (see section 2.7.2) can arise. Checking will eliminate the checking resources in the way originally intended.

A second advantage of this approach is its potential for eliminating counter-cyclic operations. The “links,“ i.e. new immediate dominance relationships, can be created bottom-up in strictly cyclic fashion. A Spell-Out mechanism sensitive to the strength property of features checked will then decide on which “link“ will be crucial in determining a constituent's linear position. Thus, taking (127) to underlie a simple English SVO clause, that Spell-Out mechanism will have to ignore that the direct object is “weakly linked“ to Spec,AgrOP. VO order is then derivable.

This approach to weak movement targets, also called “single pass syntax“ by Stabler (1996) and “spell-out at the interface“ by Groat&O'Neil (1996), removes the second factor responsible for the “explosion problem,“ namely, countercyclicity (cf. H3 of section 2.5.2). I will therefore formulate hypothesis 5.

Hypothesis 5: The proper treatment of (unbounded) dependencies in minimalist syntax requires appeal to “multiconstituency“/“multidominance“ (MC/MD) (=H5)

Appeal to multiconstituency/multidominance as such is neither spectacular nor new.[223] What is interesting about it is that it can be considered quite a natural solution to the specific problem situation current minimalist syntax creates. Section 3 will be devoted to exploring this possibility.

2.7.4 Feature Movement, Satisfy, and Agree (An Excursus)

Chomsky (1995a, section 4.4.4) introduces a further radicalization of movement theory. Given that movement is driven by feature checking requirements, the minimal object to undergo movement would seem to be not full-fledged terms but just their (formal) features, i.e. FFORM.[224] Movement is therefore taken to be "really" just feature movement, called "Move F." However, pure Move F only affects the post Spell-Out computation toward λ. π-visible movement is assumed to pied-pipe at least the minimal term containing FFORM, due to "properties of the phonological component" (Chomsky 1995a, p.263). Independent evidence for a more abstract approach to covert movement can be found in the domain of binding and scope, as (128) (cf. Brody 1995, p.133) and (129) (cf. Lasnik 1999a, p.4) illustrate.

(128) a. John wondered [which picture of himself]$_i$ Mary saw t_i
b. * John wondered when Mary saw [which picture of himself]

(129) a. Many students aren't in the room
b. There aren't many students in the room

If covert movement were to affect the whole *wh*-in-situ constituent in (128b), the binding domain for the anaphor *himself* would be expected to be extended to the matrix clause, as in (128a). Likewise, if full-fledged "associate raising" of *many students* took place after Spell-Out in (129b), it would be unclear why the indefinite can't scope over negation here, which it can in (129a). Raising just FFORM, i.e. Move F, should preclude these unwelcome possibilities. Wrt scope taking, the issue is actually less clear, given that QR is implemented as covert feature raising as well (Chomsky, 1995a, p.377).[225]

There are a number of open questions concerning the details of Move F. First of all, the internal structure of syntactic terminals hasn't been made explicit. Thus, it is unclear whether formal features would count as terms and thus be considered syntactic constituents (cf. Chomsky 1995a, p.383fn.27). Given that moved FFORM objects are adjoined to syntactic heads (ibid., p.360f), the head of an FFORM chain does count as a term (cf. section 2.6.2 and 2.6.3). This raises questions about chain uniformity (cf.

---

[223] It has been defended wholly or in part by Engdahl (1986), Gazdar et al. (1985), Pollard&Sag (1987, 1994), and Blevins (1990) among many others.

[224] For an earlier discussion of this possibility, see Muysken&vanRiemsdijk (1986, p.24).

[225] Likewise, attempts have been made by Pesetsky (2000) and Wilder (1997) to show that covert movement must affect full-fledged terms.

2.7.1). Likewise, successive-cyclic X°-movement needs some reconsideration. Assume for example that X° adjoins to Y° before Spell-Out, forming $[_{Y°}$ X° Y°]. If Y°-to-Z° is forced after Spell-Out, this should be feature movement and thus affect FFORM of Y°. However, features of X° are typically also relevant for checking against Z° (or items in Spec,ZP). Under common assumptions, these would have been pied-piped by movement of Y°. Under Move F, however, this is not guaranteed due to the fact that FFORM of X° and FFORM of Y° do not form a constituent. The latter issue would actually be sidestepped if syntactic terminals were identified with their FFORM-bundles right from the start (cf. section 2.6.1 and Jackendoff 1997).

Finally, note that, since Move F continues to involve copying and counter-cyclic operations,[226] arguments concerning the "resource paradox" and the "explosion problem" (section 2.7.2) remain unaffected by this modification of movement theory.
An even more abstract perspective on syntactic movement operations is developed in Frampton&Gutmann (1999) and Chomsky (2000, 2001). The gist of this approach is clearly brought out by the following quote.

> "This extension is based on a shift in point of view about the relation of feature checking and movement. Chomsky's early view was that movement was undertaken in order to bring heads into a local relationship so that checking could be carried out. The checking itself was assumed to take place under strict locality. Particularly with Chomsky's shift from viewing it as Attract, there is a peculiar twist to this logic. Why should strict locality be necessary for checking? Locality is generally required to establish syntactic relationships. But in the Attract framework, the relationship must be first recognized long-distance by feature matching before any local relationship is established. The local relationship which results from movement is after the fact of establishing the feature matching relationship. Rather than assuming that movement is required for checking to take place, we assume that movement is a consequence of checking" (Frampton&Gutmann 1999, p.11).

Indeed, viewing movement as induced by a functor ("attractor") searching for an appropriate argument in its c-command domain (cf. Chomsky 1995a, section 4.5.6) paves the way for allowing long-distance checking, contrary to earlier assumptions (cf. section 2.7.2). Frampton&Gutmann (1999) restate movement theory in terms of a complex transformation called "Satisfy" ("pivot" = the functor in target position).

---

[226] Roberts (1998) argues for applying Move F before Spell-Out as well. Move F vs. "Move category" is taken to correspond to the weak vs. strong distinction among features and the landing sites they control. Allowing overt Move F to be interspersed with full-fledged term (category) movement voids the need for counter-cyclic operations and thus potentially steers clear of the "explosion problem." For further discussion, see Alexiadou et al. (forthcoming). Move F has been explicitly abandoned in Chomsky (2000, p.119).

(130) Satisfy
a. location of a head Y which the pivot X recognizes;
b. carrying out feature checking between X and Y; and
c. displaying the feature checking structurally

“Recognition“ requires the features of X and Y to match (in relevant respects). Crucially, checking applies to a potential “voyager“ before movement. Schematically, a transition from (131a) to (131b) is possible.

(131) a. [. . . *F . . . [. . . F* . . .]]
b. [. . . F . . . [. . . F . . .]]

The compounding of (130a) and (130b) corresponds to the operation “Agree,“ introduced in Chomsky (2000, p.101).[227]

The third step of “Satisfy,“ consists of the decision in what way checking should be “displayed.“ This roughly boils down to the choice between moving (a copy of) X into the projection of Y and leaving X in situ.[228] This choice, of course, is largely determinable on the basis of strong vs. weak features, only the former requiring displacement. It is only this displacement option which is called “Move“ in Chomsky (2000, p.101).

One of the obvious advantages of this system lies in its potential for dispensing with counter-cyclic operations. Abstract checking of weak features can be interspersed in a cyclic bottom-up fashion. Thus, no “explosion problem“ seems to arise under such an approach.

In addition, it looks as if the “resource paradox“ had likewise evaporated. If checking precedes copying, no unchecked feature will ever be copied and checking resources won't be proliferated. Schematically, overt movement would be a transition from state (131b) to (132).

(132) [. . . [F] F . . . [. . . ($_{COPY}$[F]) . . .]]

However, current analyses of successive cyclic movement and partial movement seem to stand in the way of such rapid and complete success. Take raising constructions. Whatever drives the transition form (133a) to (133b) will not lead to checking of the uninterpretable Case feature of DP *John*, abbreviated as C*.

(133) a. [to be likely [John^{C*} to fall asleep]]
b. [John^{C*} to be likely [($_{COPY}$ John^{C*}) to fall asleep]]

[227] For exegetic reason, I discuss the core issues in terms of Frampton&Gutmann (1999), assuming that they can be rephrased in terms of Chomsky's (2000, 2001) system.
[228] I ignore the more abstract variants of “display“ discussed in Frampton&Gutmann (1999).

Under the copy theory of movement, checking resources will have been copied. Thus, the "resource paradox" has been regained.[229] Next consider an ECM construction like (134).

(134) We believe [John to have repeatedly been criticized by his boss]

These could be considered cases of "partial" movement in the sense that the π-visible position of *John* is not immediately contained in the projection of the "Case-assigner," the latter being somewhere in the matrix clause. Thus, in order to generate (134) we need one application of Satisfy without Case checking but leading to overt "display" followed by an application of Satisfy with Case checking but without overt "display." The first step, however, will have led to the copying of an unchecked uninterpretable Case feature, which is left inside the embedded VP. This copied unchecked Case feature will, given standard assumptions, be unaffected by the second application of Satisfy, which establishes a checking relation between the Case-assigner and the head of the A-chain. Thus, we only get a transition from (135a) to (135b). [*C is (part of) the Case-assigner and C* the unchecked Case-feature.]

(135) a. We *C believe [John^{C*} to have repeatedly been criticized ($_{COPY}$ John^{C*}) by his boss]
b. We C believe [John^{C} to have repeatedly been criticized ($_{COPY}$ John^{C*}) by his boss]

Next, suppose we apply remnant VP-topicalization to (135b), as illustrated below.

(136) [criticized ($_{COPY}$ John^{C*}) by his boss], we believe [John^{C} to have repeatedly been ($_{COPY}$ [criticized ($_{COPY}$ John^{C*}) by his boss])]

Suppose in addition that (136) were made a constituent undergoing A-movement in a larger piece of structure.

---

229 Nothing hinges on which constructions are actually chosen to establish this point. What is crucial is that chain formation must take place at a point where (complete) checking is not possible. In Chomsky (2000, 2001) this typically involves satisfaction of the "Extended Projection Principle" (EPP) in the projection of a "deficient" I°, i.e. the head of raising infinitivals, which tends to be taken to provide an intermediate landing site for A-movement into its specifier. Chomsky (2001, p.8f) mentions an alternative that would not allow EPP-checking in such a position. More importantly, the necessity of creating A-chains has been fundamentally questioned in Hornstein (1998, 2000), and Lasnik (1999b). On the other hand, Chomsky (2000, p.128; 2001, p.47fn.49) tentatively assumes that *WH*-phrases bear uninterpretable *WH*-features. Such features, of course, belong to the kind of checking resources this section has been concerned with. Thus, as long as *WH*-movement proceeds successive-cyclically, leaving copies in intermediate positions, another variant of the "resource paradox" arises. Also, every kind of "anti-freezing" effect would have to be reviewed. See papers in Alexiadou et al. (forthcoming) for potential examples.

(137) a. [That [criticized ($_{COPY}$ John^{C*}) by his boss], we believe [John^{C}
to have repeatedly been ($_{COPY}$ [criticized ($_{COPY}$ John^{C*}) by his boss])]]
seems t'''' to be likely t''' to appear t'' to be believed t' to sound t
ridiculous

b. t = [That [criticized ($_{COPY}$ John^{C*}) by his boss], we believe [John^{C}
to have repeatedly been ($_{COPY}$ [criticized ($_{COPY}$ John^{C*}) by his boss])]]

Again, an unresolved checking task, to be dealt with by non-standard means, has been proliferated into copies of copies. Without extra assumptions, another type of "explosion problem," this one independent of counter-cyclic movement, may well arise.

I conclude that hypotheses 3 and 5 are defensible even in the light of more recent developments. Countercyclicity has to be avoided and the power of copying must be restrained. One way of doing the latter, has been sketched in 2.7.3 and will be explored further in section 3.

2.8 (Restricted) Economy

The previous sections have illustrated the core syntactic part of what it means that C_{HL} enumerates the set of well-formed or "convergent" expressions. Merge and Move map selections of members from LEX into pairs of representations, <π,λ>, the latter each obeying the Principle of Full Interpretation (FI)(cf. section 2.2).

Now, there has been another strand of ideas explored in the minimalist program, according to which the above mapping can be freed from various stipulative conditions if economy principles are recognized as additional filters. Technically, removing conditions on the mapping from LEX to <π,λ>-pairs enlarges the set of "convergent" expressions. The claim, then, is that some kind of natural competition can cut this set down in a conceptually and empirically adequate way to a subset of "admissible" expressions, that is, to the set of expressions the grammar ultimately recognizes as "grammatical" (cf. section 2.2).

However, due to the successive refinements documented in Chomsky (1995a), the domain of application for genuinely competitive principles has been dwindling, a fact which gives me the opportunity to skip the thorny technical and conceptual issues involved and concentrate on the clearest remaining case instead.[230]

---

[230] Most notably, the "Minimal Link Condition," originally taken to be a transderivational implementation of Rizzi's (1990) theory of "Relativized Minimality," has been reanalyzed as a condition on the operation Move, definable over a single input phrase-marker. Nakamura (1998) provides a very interesting application of "Shortest Path," a transderivational variant of the MLC, to extraction in Tagalog. However, his modification of the notion of reference set seems to have the unwelcome consequence that, for example, the optionality of WH-movement vs. WH-in situ in French constituent questions can no longer be derived. For discussion of technical and conceptual matters concerning economy principles see Grewendorf (1995), Kitahara (1995, 1997), Nunes (1995), Büring (1996), Stabler (1996), Collins (1997, 2001), Johnson&Lappin (1997, 1999), Kolb

A basic understanding of this core case requires a brief retracing of how "guidelines [which] have a kind of "least effort" flavor to them" are elevated to "actual principles of language" (Chomsky 1995a, p.130).

Take for example the analysis of X°-movement in French. The following contrast has been explained by counting the number of operations required to derive each structure.

(138) a. Jean embrasse souvent Marie
John kisses often Mary
b. * Jean souvent embrasse Marie

Simplifying matters radically,[231] we can say that (138a) blocks (138b) because the derivation of (138a) employs one step less than the derivation of (138b). Both involve V°-to-I° movement, overt in (138a) but covert in (138b). Under the then current analysis, (138b) would require an additional, and therefore crucially superfluous, I°-to-V° lowering operation before Spell-Out. This mechanism has later been abandoned (cf. section 2.4), so the example is purely illustrative. It is immediately clear, although this was not discussed at the time when (138) was first analyzed in terms of economy principles, that blocking effects cannot just be a global numerical property. Thus, (139a), which arguably involves fewer derivational steps, certainly does not block (139b).

(139) a. Humpty Dumpty is snoring
b. I can hear that Humpty Dumpty is snoring

Trivial though this point may appear at first sight, it is essential that, for transderivational economy priniciples to function properly, "candidate-" or "reference sets" be determined among which economy principles can select optimal competitors.[232] As a first approximation, Chomsky (1995a, p.227) takes economy principles to consider only those competitors that go back to identical numerations, $\mathbb{N}$. Thus, (140a) and (140b) can peacefully coexist, because they derive from different numerations.

(140) a. A unicorn is in the garden
b. There is a unicorn in the garden

The same holds for the pair in (139), of course. Now, the crucial application of a genuine economy principle in minimalist syntax concerns the following contrast.[233]

(1997a), Epstein et al. (1998), Uriagereka (1998), Frampton&Gutmann (1999), Lasnik (1999a), as well as papers in Abraham et al. (eds.) (1996), Wilder et al. (eds.)(1997), Barbosa et al. (eds.)(1998), and Müller&Sternefeld (eds.)(2001).

[231] For the pre-minimalist analysis see Pollock (1989). See Grewendorf (1990) and Lasnik (1999a) for detailed discussion, including mention of some recalcitrant problems with this analysis.

[232] Cf. Sternefeld (1997) for thorough discussion.

[233] Cf. Chomsky (1995a, p.344ff).

(141) a. There$_i$ seems t_i to be someone in the room
b. *There seems someone$_i$ to be t_i in the room

A superficial look predicts that the same number of operations has to be applied in both cases. Indeed, (141) is supposed to vindicate a much subtler interpretation of competition along with the bottom-up nature of derivations. A general premise for this is the assumption that Move operations are costlier than Merge.

(142) Cost of Operations[234]
Move > Merge

Secondly, economy is computed "locally," that is, (142) has a say only if a choice point arises, such that each of a number of possible continuations would lead to convergence. This has been made more explicit by Frampton (1997, p.37), whose definition I will take over in a slightly adapted way (cf. section 2.6.2)

(143) Suppose $D = (\Sigma_1, \ldots, \Sigma_n)$ is a derivation. Then the derivational stage Σ is called a *well-formed continuation* (wfc) of D if there are derivational stages T_i, $(1 \leq i \leq m)$, such that $(\Sigma_1, \ldots, \Sigma_n, \Sigma, T_1, \ldots, T_m)$ is a well-formed derivation (wfd).

(144) A wfd $D = (\Sigma_1, \ldots, \Sigma_n)$ is called *optimal* if for all i $(1< i \leq n)$, Σ_i is an optimal wfc for $D = (\Sigma_1, \ldots, \Sigma_{i-1})$

(145) In a set $\mathbb{T} = \{ T \mid T \text{ is a wfc for } D = (\Sigma_1, \ldots, \Sigma_n) \}$ T_i is *optimal* if there is no $T_j \in \mathbb{T}$, such that the operation that produces T_j from Σ_n is lower on the cost scale (142) than the one producing T_i from Σ_n

Thus, optimal well-formed derivations, i.e. the ones obeying economy principles over and above convergence conditions, can be used to define the above-mentioned set of "admissible" expressions. Returning to the contrast in (141), we get the following account. At the stage where (146) has been reached, the specifier of IP must be filled.

(146) [$_{I'}$ to be someone in the room]

Either *there* is inserted by Merge (147a), or *someone* is raised by Move (147b).

(147) a. [there to be someone in the room]
b. [someone$_i$ to be t_i in the room]

[234] Cf. Chomsky (1995a, p.226). In terms of subroutines this hierarchy makes sense, given that Merge is considered a proper part of Move, though definitions require some sharpening (cf. section 2.6.2). Intuitively, one could argue that the numeration has to be emptied as fast as possible to reach the end of the derivation.

Both are wfcs of (146), leading to (141a) and (141b), respectively. However, only (147a) involves an *optimal* wfc of the structure in (146), given that it results from an application of Merge as opposed to Move (147b), hence being chosen in accordance with (142)-(146). Given this local decision, no further option arises when it comes to creating the matrix specifier, i.e., when it comes to completing (148).

(148) [$_{I'}$ seems [there to be someone in the room]]

There, being the closest eligible DP, undergoes raising, which yields (141a). The ungrammatical (141b) is properly ruled out by this local bottom-up computation in terms of economy.

Yet, this restricted version of economy should not be adopted, since it incorrectly predicts (149a) to block (149b).[235]

(149) a. [A rumor [$_{CP}$ that there was a unicorn in the garden]]$_i$ was t_i in the air
b. There was [a rumor [$_{CP}$ that a unicorn was t_i in the garden]] in the air

The logic of embedding requires the CP-complement of *rumor* to be completed first, giving rise to a choice point on filling the IP specifier of that CP. In parallel with what was posited above, *there* has to be inserted by an application of Merge and (150a) is transformed into (150b).

(150) a. [$_{I'}$ was a unicorn in the garden]
b. [$_{IP}$ there was a unicorn in the garden]

On creating the specifier of the matrix IP, once again no option is left. *There* having been taken out of the numeration, movement of the complex DP is all that can be done, turning (151a) into (151b).

(151) a. [$_{I'}$ was [a rumor [$_{CP}$ that there was a unicorn in the garden]] in the air]
b. [$_{IP}$ [A rumor [$_{CP}$ that there was a unicorn in the garden]]$_i$ was t_i in the air]

Of course, the reverse order of Move and Merge, required to derive (149b), is not available under the assumptions made, much to the detriment of the entire account.[236]

---

[235] The following has already been published in Wilder&Gärtner (1997). See also Büring (1996) and Johnson&Lappin (1999).

[236] For an alternative analysis see Frampton (1997) and Lasnik (1995). For a revision of assumptions, including an appeal to cyclic optimization, see Chomsky (2000). The idea that Merge should preempt Move may be further challenged by considering the case of WH-expletives in German. Thus, according to "direct dependency" approaches (cf. papers in Lutz et al. (eds.)(2000)), *was* in (i) is an expletive WH-scope marker associating with the lower WH-phrase *wen* (indicated by co-superscription.)

(i) Was^{i} meinst Du, [wen^{i}]$_j$ (dass) wir t_j einladen sollen?
What mean you who (that) we invite should

In sum, given (i) the lack of examples for transderivational economy principles which are both convincing and empirically sound, (ii) the nontrivial task of implementing putative economy principles, as the sketchy formulations in (142)-(145) indicate, and (iii) the fact that transderivational principles do not belong to the simple steps of theory-design appealed to in the minimalist quest for an answer to Q1 (cf. section 1.1), I will defend hypothesis 6.

Hypothesis 6: There are no economy principles operating in minimalist syntax (=H6)[237][238]

"Who do you think we should invite?"

As can be gathered from the translation, (i) is equivalent to the direct constituent question in (ii), lacking *was*.

(ii) [Wen]$_j$ meinst Du, t_j (dass) wir t_j einladen sollen?

Importantly, however, *was* must not be inserted "as soon as possible." Instead, the associated WH-phrase has to perform at least one movement step. This is shown by the unacceptability of (iii).

(iii) * [Was^i]$_j$ meinst Du, t_j (dass) wir wen einladen sollen?

This is unexpected if there is an economy principle like (142), forcing Merge to block Move in such cases. Given (142), we should expect (iii) to block (i). For an alternative, "indirect dependency" approach, see Staudacher (2000).

237 I've already indicated in section 2.7 that if Stabler (1996) is correct in assuming that "economy conditions may have natural implementations in the performance model," we would, provided convincing examples for such principles can be found, have a strong vindication of "Frampton's conjecture" introduced in section 1.1 above. It is probably no surprise that such a claim should come from the area of parsing, which is generally much more at ease with computational techniques. Let me, therefore, in elaboration on section 1.1, offer some speculation as to why "economy principles" might have come to the attention of competence linguists. This has to do with some suggestive analogies from computer science and cognitive science. "[. . .] it is characteristic of the organization of general purpose digital computers that they do not communicate in the languages in which they compute and they do not compute in the languages in which they communicate. The usual situation is that information gets into and out of the computational code via the operation of compiling systems which are, in effect, translation algorithms for the programing languages that the machine 'understands.' The present point is that, if the view of communication I have been commending is true, then these remarks hold, in some detail, for the mechanisms whereby human beings exchange information via natural languages. To all intents and purposes, such mechanisms constitute 'compilers' which allow the speaker/hearer to translate from formulae in the computational code to wave forms and back again" (Fodor 1975, p.116). Now, compiling consists of three core procedures, lexical analysis, syntax analysis, and code generation. Additionally, "[m]any compilers have a fourth phase, called *optimization*, which follows code generation. Its purpose is to make the object program smaller or quicker to execute (by techniques such as detection and elimination of redundant statements, making loop bodies as short as possible, and using registers instead of memory cells whenever feasible). The improvements are often only marginal, and are gained at the expense of additional complexity in the compiler and extra time during compilation. The extent to which it is used should therefore be governed by how often the object program is to be executed" (Goldschlager&Lister 1988, p.199).

Now, if, for the sake of argument, the operations of C_{HL} are considered part of a compilation process in the sense of Fodor, the question of optimization might also arise. However, if the

I take this cautious assumption to be vindicated by Collins (2001, p.61), who concluded that

> "[w]hat can be said with certainty is that our understanding of economy at this point is minimal."[239]

2.9 Summary

Let me now briefly summarize this fairly lengthy survey of minimalist syntax. The minimalist program, as presented in Chomsky (1995a), constitutes a further step in exploring how best to attain the "goals of linguistic theory" (cf. section 2.1 above), which implicitly continue to consist in providing

(152) a. an enumeration of the class $s_1, s_2, \ldots$ of possible sentences
b. an enumeration of the class $SD_1, SD_2, \ldots$ of possible structural descriptions
c. an enumeration of the class $G_1, G_2, \ldots$ of possible generative grammars
d. specification of a function *f* such that $SD_{f(i,j)}$ is the structural description assigned to sentence s_i by grammar G_j, for arbitrary i, j.
(Chomsky 1965, p.31)

In response to the overarching minimalist concern

Q1: How "perfect" is language?

a reduction of grammatical components as well as a refinement of grammatical operations and objects is proposed. As to the former, the computational system of human language, C_{HL}, is taken to be the minimal device capable of relating sound and meaning. This is considered to justify a two-level approach, postulating a morpho-

individual signals or "formulae," to use Fodor's expression, are taken as the object programs, it is highly unlikely that optimization is worth its while, given that each "formula" is used only once. Rather the compiler itself, assumed to be part of the human innate endowment, might be expected to have been optimized under evolution. I'm afraid, attaining greater precision here would require answers to the difficult questions section 1.1 raises, so I won't elaborate on this.

238 Collins (1997) argues for a "local" interpretation of economy principles. Thus transition Σ_i,Σ_j is blocked by transition Σ_i,Σ_k if the latter involves operations that are more "minimal." No appeal to successors of Σ_j/Σ_k is made and eventual convergence is not taken into account. As far as I understand that work, however, its main empirical applications only show that global economy isn't desirable. Its main technical result, namely, the derivation of the fact that Merge must be binary rather than *n*-ary (n≠2) comes at the cost of stipulating that unary Merge is ruled out (Collins 1997, p.81). For comprehensive discussion of "local economy," see Johnson&Lappin (1997, 1999).

239 There may be an issue whether to call competition-less principles "economy principles." For my failure to do so, I refer the reader to Schoemaker (1991), where it is shown to what extent "economy" can be in the eye of the beholder.

phonological component, PF, and a logico-syntactic component, LF, interfacing via designated representations π and λ with the articulatory-perceptual, (A-P), and the conceptual-intentional, (C-I), systems of the mind/brain (cf. sections 2.2, 2.3, and 2.4). Conditions on representations can only apply at π and λ, most important among them the Principle of Full Interpretation (FI) (cf. section 2.2).

(153) The Principle of Full Interpretation (FI)
A representation R *satisfies FI* iff R does not contain uninterpretable ("superfluous") elements

FI assumes its core role in the definition of well-formedness, called "convergence" in minimalist theory (cf. 2.2).

(154) "A derivation *converges at* one of the interface levels if it yields a representation satisfying FI at that level, and *converges* if it converges at both interface levels, PF and LF; otherwise it *crashes*."

Most importantly, it follows from (154) that π and λ have to be free of so-called "uninterpretable" features in order to be well-formed. Apart from nominal case and verbal agreement, uninterpretable features are largely confined to properties of functional heads. Together they constitute "checking resources," where "checking" means the elimination of uninterpretable features in certain configurations, comparable to cancelation operations in categorial grammar. Given (154), syntactic derivations are driven by the need to eliminate uninterpretable features.

Assuming checking configurations to be local (i.e. head-specifier relations and head-head adjunction structures), adequate syntactic description of "constructions" like WH-movement into Spec,CP, verb second movement to C°, or subject-raising into Spec,IP, depends on specification of the functional heads involved (C°, I° . . .). Cross-linguistic variation follows from whether or not operations apply before or after PF "branches off" from syntax, i.e. before or after "Spell-Out" makes syntactic information available to PF, ultimately "instructing" the A-P systems. Thus, either checkable features are taken to be "strong" and thus "π-visible," forcing "overt" elimination (before Spell-Out). Or else they are "weak" and thus not π-visible, allowing for "covert" elimination (after Spell-Out) (cf. section 2.4).

I have argued that interaction with checking resources exhausts the scope of FI. Thus, interpretability remains a largely syntax-internal notion, the interface-oriented set-up of minimalism being merely strategic. Chief emphasis is put on how to constrain the inner workings of C_{HL}. In that respect, the minimalist program can be narrowed down to a theory of minimalist *syntax* for the following reason. If successful, the minimalist program shows that what originally was called the level of phrase structure and the transformational level (Chomsky 1975/1955[1]) can be unified (cf. sections 2.3 and 2.4), a claim quite regularly made outside the Chomskyan branch of generative linguistics. How directly syntax affects the A-P and C-I systems would depend on genuinely morpho-phonological and semantic argumentation, conspicuously absent from Chomsky (1995a) except for the caveat that at least wrt the nature of operations, LF and PF are not uniform (Chomsky 1995a, p.227).[240] Even granting the programatic nature of minimalism, it is safe to first of all judge its success syntax-internally, construing the program as basically a version of "narrow syntax" (cf. section 2.2 and 3.1 below).[241]

Reducing C_{HL} to PF and LF, minimalist syntax dispenses with the non-interface levels of syntactic representation, that is, D-structure and S-structure (cf. section 2.4). It thus becomes an open issue whether principles like the Theta Criterion should be carried over from GB, the predecessor of minimalism.

(155) Theta Criterion
Each argument bears one and only one Θ-role, and each Θ-role is assigned to one and only one argument (Chomsky 1981, p.36).

Although the discussion in Chomsky (1995a) is somewhat inconclusive, I assume that (155) may still be necessary for minimalist syntax, *under certain conditions* (see sections 3.4.2 and 3.5, below). Thus, I postulate the (conditional) hypothesis 2.

H2: Under certain conditions, (part of) the theta criterion has to be implemented as a syntactic well-formedness constraint

The core part of this study then addresses the nature of operations and objects involved in minimalist syntactic computations. Along with the elimination of D-structure and S-structure, the separation of phrase structure component and transformational component is given up. Syntactic constituent structures are built "bottom-up," accessing a pool of lexical resources called "numeration," $\mathbb{N}$, directly.

Two operations are involved in this derivational process. First, there are *binary* or "generalized" transformations, referred to as "Merge" from Chomsky (1995a, chapter 4) on. These join two structures into a compound structure, as schematically illustrated in (156).

(156) BT(α,β)
$\alpha \quad \beta \rightarrow (\alpha, \beta)$

Secondly, there are *singulary* ("movement") transformations, referred to as "Move" from Chomsky (1995a, chapter 4) on. These apply to a complex structure, lift a substructure out of it, and join the two elements into a compound structure that preserves a "trace" or "copy" of the original substructure. This is schematically illustrated in (157).

[240] It remains to be shown how directly works like Halle&Marantz (1993) and Reinhart (1995) can be integrated with minimalist syntax, for a broader, interface-oriented perspective to be vindicated.
[241] Cf. Chomsky (2000, p.100).

(157) ST(($_\alpha$ β))
($_\alpha$ β) → (β, ($_\alpha$ ($_{\text{COPY}}$ β)))

Sections 2.5 and 2.6 deal with two different ways of looking at the objects manipulated here. The first one goes back to Chomsky (1995a, chapter 3), which considers the objects transformed in (156)/(157) as phrase markers or constituent structure trees. Due to its appeal to X-bar theory, this system inherits a number of unresolved questions from its GB-predecessor.

The second one is defended in Chomsky (1995a, chapter 4) under the label "bare phrase structure." This time, the objects transformed in (156)/(157) are considered sets. Starting point for this attempt at reduction is an appeal to so-called "inclusiveness" (cf. section 2.6.1) (Chomsky 1995a, p.228).

(158) The Principle of Inclusiveness
Any structure formed by the computation (in particular, π and λ) is constituted of elements already present in the lexical items selected for N; no new objects are added in the course of computation apart from rearrangements of lexical properties.

Inclusiveness bans the use of bar-level markers, thus rendering classical X-bar theory obsolete, its content being replaced by means of contextual definition (cf. 2.6.4, 2.7.1) (Chomsky 1995a, p.242).

(159) "A category that does not project any further is a maximal projection XP, and one that is not a projection at all is a minimal projection X^{min}, any other is an X', invisible at the interface and for computation."

Likewise, indices, appealed to in GB to determine binding relationships and chain membership, are taken to be unavailable. Some indexation, however, is made use of in order to individuate "occurrences" of identical types throughout the computation. This will play an important role in section 3.3 below.

A closer look at BT and ST reveals that the former could by a subroutine of the latter. Thus β and ($_\alpha$ ($_{\text{COPY}}$ β)) in (157) undergo set-formation, i.e. "Merge." As a consequence, one of the peculiar properties of BT, namely, the "growth effect," can be turned into a general principle. In its most radical form, this type of "cyclicity," discussed in Chomsky (1995a, chapter 3) under the term "extension condition" (cf. 2.5.1), requires every syntactic operation to create a new "root" (node). Crucially, applications of ST to objects already embedded in larger structures would be ruled out, as schematically illustrated in (160).

(160) ST(($_\alpha$ β))
* ($_\gamma$ ($_\alpha$ β)) → ($_\gamma$ (β, ($_\alpha$ ($_{\text{COPY}}$ β))))

This motivates my hypothesis 3.

H3: There are no counter-cyclic syntactic operations

Section 2.5.1 discusses and dismisses empirical arguments against H3, involving the interaction of binding theory, reconstruction, and adjunction structures. Formal obstacles to the implementation of H3, arising from covert movement and X°-movement, are dealt with below in sections 3.3. and 3.4.2, respectively (cf. section 2.7.3).

Chomsky (1995a) pursues a weaker version of H3, valid only for Merge. This is accompanied by a fundamental distinction of "syntactic objects" (≈ phrase markers) and "terms" (≈ sub-phrase markers), the latter defined relative to the former. This distinction, implying the existence of two Merge operations, stands in the way of integrating Merge and Move. I consider these properties questionable (cf. section 2.6.2).

Q3: (i) Is there any need for both syntactic objects and terms?
(ii) Should there be more than one type of Merge?

The answer given to both parts of Q3 in section 3 is negative.

The issues of phrase structural status and interpretability, as governed by FI meet in the notion of "visibility" (cf. section 2.6.4). According to this notion, objects are "visible" and thus accessible to operations of C_{HL} only if they are interpretable. Although this may in principle yield welcome results concerning intermediate projections, the overall idea is highly problematic for at least three reasons. First, it is unclear how to prevent invisibility of intermediate "projections" from causing invisibility of the entire substructure "dominated" by that projection. This is due to the set-theoretic make-up of objects, identifiable only by their members, given the absence of any node/decoration distinction. Secondly, semantic evidence for the interpretability of structures considered intermediate projections in syntax is not hard to come by. Thirdly, extending the conspiracy of interpretability and visibility to features would render the driving force of syntactic derivations, i.e. "uninterpretable" features, ineffective right from the start. This motivates my hypothesis 4.

H4: Invisibility of syntactic elements for C_{HL} has either to be avoided or it must be stipulated and properly implemented

The validity of H4 is further vindicated if "chains" are considered (cf. section 2.7.1). Thus, the pair of terms < β, ($_{COPY}$ β) > resulting from Move in (157) is taken to constitute a higher-order object called "chain." Originally, chains were taken to constitute the only "legitimate," and thus "interpretable," objects at λ (cf. Chomsky 1995a, p.154). The inventory of such objects is given in (161).

(161) Legitimate LF Elements (each a chain)
a. Arguments: each element is in an A-position, α_1 Case-marked and α_n Θ-marked
b. Adjuncts: each element is in an A'-position
c. Lexical elements: each element is in an X° position
d. Predicates, possibly predicate chains if there is predicate raising
e. Operator-variable constructions (α,β), where α is in an A'-position and β heads a legitimate uniform chain

However, given that chain formation is a result of Move, the input arguments of that operation, not being chains themselves (unless some kind of "type-lifting" is assumed) should not be accessible to C_{HL}. If the uninterpretability/invisibility conspiracy is understood in its strictest sense, non-chains are simply invisible. Again an unwelcome result.

An additional question concerns the effect of copying on checking resources. The latter come in pairs to be canceled in local configurations, and thus driving movement operations, i.e. transitions like the one in (162).

(162) a. [. . . *F . . . [. . . F* . . .]]
b. [. . . [F*]$_i$ *F . . . [. . . t_i . . .]]
c. [. . . [F]$_i$ F . . . [. . . t_i . . .]]

If, however, the "trace" of Move, t_i, is a full copy of the moved constituent, ($_{COPY}$ F*), uninterpretable features should always survive in non-local position and subsequently cause FI-violations. This can be called a "resource paradox" (cf. section 2.7.2).

There are basically three approaches to this problem. First, one can non-locally check off features before applying Move. This is suggested by Frampton&Gutmann (1999) as well as Chomsky (2000, 2001). Thus, (162a) would be followed by (163) before Move applies.

(163) a. [. . . F . . . [. . . F . . .]]

This strategy has not yet been shown to cover the desired range of cases and may thus still be insufficient (cf. 2.7.4). Secondly, one may take checking to apply to entire chains, as proposed by Chomsky (1995a, p.381fn.12). This requires a worked out format for chain structures. Also, care must be taken to avoid countercyclicity and the concomitant postponement of checking. Otherwise, checking resources may be further multiplied without direct chain relations being established. This is schematically illustrated in (164).

(164) [[G* [F*]] *G [. . . *F . . . [. . . ($_{COPY}$ [G* [F*]]) . . .]]

I have called this the "explosion problem" and take it to be further motivation for H3.

The third approach, advocated here, disallows copying and assumes a conception of structures that makes it possible for one and the same element to be an immediate constituent of more than one larger constituent (cf. section 2.7.3). In graph-theoretic terms, the difference between this and the copy-approach is given in (165).

(165) a. b.

It is clear why no resource paradox can arise if structures like (165b) replace the ones in (165a). This gives rise to the central hypothesis of this study, H5.

H5: The proper treatment of (unbounded) dependencies in minimalist syntax requires appeal to “multiconstituency“/“multidominance“ (MC/MD)

Section 3 will be devoted to showing how this kind of structures can be built up in compliance with H3 as well.

Section 2.8 confronts one of the major arguments for non-trivial, transderivational economy principles in minimalist syntax with a serious technical problem. I take this to be sufficient grounds for assuming hypothesis 6.

H6: There are no economy principles operating in minimalist syntax

3 The Syntax of Multiconstituency and Multidominance

Wie die Kuh über's Kirchendach!
(Fitzcarraldo)

Having completed my survey of the minimalist program, I'm now in a position to propose a modified version of minimalist syntax called "the syntax of multiconstituency and multidominance," or, in short, "MC/MD-syntax."

Setting up the system will require two preparatory steps. First, section 3.1 "narrows down" the scope of the theory to minimalist *syntax*. Crucially, the role of interfaces will be minimized, there being at this point no substantial evidence for principles governing minimalist syntactic processes from the outside. In particular the interpretation-induced notion of "invisibility" will be dispensed with, in fulfillment of hypothesis 4 (cf. section 2.6.4, 2.7.1). Section 3.2, secondly, surveys "varieties of multiconstituency and multidominance" from a graph-theoretic perspective. This involves discussion of "phrase-linking grammar" (Peters&Ritchie 1981) and an excursus on "structure sharing," familiar from frameworks like HPSG (Pollard&Sag 1987, 1994). Also, I introduce "(rooted) MC/MD-graphs," serving as formal background against which my own system will be developed.

Then, in fulfillment of hypothesis 5, MC/MD-syntax is presented in section 3.3. The core of that system is constituted by a hybrid operation, DoIC/DoID, replacing minimalist Merge and Move. This operation defines an immediate constituency(IC-)/ immediate dominance(ID-)relation into which constituents can be multiply entered. Thus, being in multiple IC-/ID-relations replaces copies and chains of minimalist syntax. In order to capture (strictest) cyclicity and the C-Command Condition on movement, an "Ancestor Condition" on DoIC/DoID is introduced. This condition requires each application of DoIC/DoID to create a root node and to "connect" ancestors of multiply dominated constituents. Covert movement will be treated cyclically on a par with overt movement, except for keeping as a second record a so-called "weak" IC-/ID-relation. The latter allows me to prevent constituents from being "spelled-out" in covert positions.

MC/MD-syntax is developed in two variants, one based on the operation DoIC (section 3.3.1), the other on the operation DoID (section 3.3.3). As discussed in section 3.3.2, the DoIC-system, which, following minimalist syntax operates on a uniform domain called "terms," seems to involve an illegitimate use of the power of set theory. Thus, among other things, checking requires multiple substitutions, which, as I argued in section 2.7, is one of the dubious moves to salvage the copy theory of movement from the "resource problem." If, however, an appeal to set-theoretic extensionality solves the substitution problem for MC/MD-syntax, it could have solved it for minimalist syntax as well. Since I'm skeptical of such a view, I develop the DoID-system in section 3.3.3. Here, the domain of objects is node-like and partitioned into “terminals“ and “non-terminals.“ The former contain checking resources and do not enter into multidominance relations. Thus, each substitution is restricted to a single location. The non-terminals, which are free from checking resources, are entered into (multiple) ID-relations as their counterparts in the DoIC-system before. Furthermore, full individuation of nodes is brought about by indexation. Indices likewise serve the purpose for each projection of “pointing“ to the location of its checking resources, much as labeling is supposed to do in minimalist syntax.

Section 3.4 broadens the scope of MC/MD-syntax by adding an explicit C-Command definition (3.4.1) and a technique for integrating linear precedence (3.4.3). Also, four ways of dealing with X°-movement, compatible with (strictest) cyclicity, are outlined in section 3.4.2. Finally, section 3.5 confronts a number of objections to MC/MD-syntax. It is argued that these can be met if explicit PF mechanisms are assumed. The latter involve, in the “worst case,“ substitution of empty terminals into “trace positions.“ Both section 3.5 and 3.4.2 discuss conditions under which hypothesis 2 would a play a role in MC/MD-syntax.

3.1 Narrow Syntax

For reasons already discussed in section 2.2, I will reduce the interface burden of syntactic computations and embed syntax in a (familiar) tripartite structure, all of which amounts to the design of a theory defining “narrow syntax.“ The following well-formedness conditions, repeated from section 2.2, will be postulated. This essentially follows Jackendoff (1997).[1]

(1) C_{HL} generates *linguistic expressions L*, which consist of triples <phon, syn, sem>, *phon* a PF-representation, *syn* a syntactic representation, and *sem* a semantic representation.

(2) A linguistic expression L (= <phon,syn,sem>) is *grammatical* iff
a. there is a derivation D that generates *syn*, and
b. *syn* can be translated into *phon* and *sem*, and
c. *phon* and *sem* are well-formed

Clearly, interfaces still play a role, namely, in the guise of translation procedures from *syn* to *phon* and *sem*. However, nothing has to be said about articulatory-perceptual and conceptual-intentional interfaces in order to judge the success of syntax. The Principle of Full Interpretation (FI) will be reformulated in the following way.

[1] Cf. Chomsky (2000, p.100). Following standard minimalist syntax, I will often use π instead of *phon*.

(3) Principle of Translatability
Syn is not translatable into *phon*/*sem* if *syn* contains unchecked O-features

“O-features,“ i.e. “offensive features,“ belong to the still to be conclusively established subset of FFORM, called “uninterpretable“ features in Chomsky (1995a, chapter 4). Over and above the features on functional heads (cf. section 2.4), Case features on nouns and agreement features on verbs are the clearest candidates for such O-features. In section 3.3.1, an addition to (3) will be made in order to capture a counterpart of the “Single Root Condition“ on trees. Other than wrt (3), interfaces have no influence on syntax proper. Thus, in fulfillment of my hypothesis 4, principles that directly induce the “invisibility“ of syntactic constructs because the latter are not “interpretable“ are off limits.

In contrast with Chomskyan minimalist syntax, Spell-Out will not be able to apply during syntactic computation (cf. section 2.4) but has to await the completion of *syn*. Thus, I advocate what Groat&O'Neil (1996) have called “Spell-Out at the LF Interface,“ or Stabler (1996, 1998) “single pass syntax.“ For reasons of parallelism, I alternatively call this procedure “translation.“ The obvious three-component structure of C_{HL} into which narrow syntax is embedded is given in (4) (< and > designate translation).

(4) PHON < SYN > SEM

Of course, translation is a nontrivial concept. Uncertainties on the PHON-branch are inherited from the minimalist system, whose rather sketchy characterization of morphophonology and the way linear order among terminal elements is induced there requires some sharpening below. As for translation into the SEM-branch, I will have nothing new to offer. Engdahl (1986) showed how to operate such a procedure on the basis of “phrase linking grammar“ (PLG). The common ground of my system and PLG will be discussed in section 3.2, to which I now turn.

3.2 Varieties of Multiconstituency and Multidominance

This section introduces some varieties of “multiconstituency“ and “multidominance“ (MC/MD) recognized in linguistics. Presentation will be neither comprehensive nor particularly deep, given the rather special use I want to make of MC/MD later.

Note, to begin with, that Chomskyan generative syntax tends to be developed with reference to structures equivalent to “trees“ as defined in (14) of section 2.3, repeated here as (5) for convenience (Partee et al. 1993, p.441f).

(5) A *constituent structure tree* is a mathematical configuration $< N, Q, D, P, L >$, where
N is a finite set, the set of nodes,
Q is a finite set, the set of labels,
D is a weak partial order in $N \times N$, the dominance relation,
P is a strict partial order in $N \times N$, the precedence relation,
L is a function from N into Q, the labeling function,
and such that the following conditions hold:
(1) $(\exists x \in N)(\forall y \in N)[<x,y> \in D]$ (Single Root Condition)
(2) $(\forall x,y \in N)[\ (<x,y> \in P \vee <y,x> \in P\ ) \leftrightarrow (\ <x,y> \notin D \wedge <y,x> \notin D\ )\ ]$ (Exclusivity Condition)
(3) $(\forall w,x,y,z \in N)[\ (<w,x> \in P \wedge <w,y> \in D \wedge <x,z> \in D\ ) \rightarrow <y,z> \in P\ ]$ (Nontangling Condition)

However, this is not a matter of principle, as the following quote illustrates.

> “Still, we can ask whether D-structures, S-structures, etc., have the properties of tree structures. Insofar as they are determined by X-bar theory, this will be the case. But there are other factors that enter into determining their properties. Reanalysis and restructuring processes, for instance, may yield phrase markers that cannot be represented as tree structures. [. . .] Furthermore, X-bar theory can be constructed so that it does not require that phrase markers have tree properties. It has occasionally been suggested that coordination might be understood in terms of union of phrase markers (in effect, three-dimensional projection), linear order being determined by a “spell-out“ rule. The assumption would be, then, that if the very same language were to be used in a medium having a dimension in addition to linear time, this “spell-out“ rule would be unnecessary. Such suggestions might be correct, and I think they merit examination. Much more radical departures from tree structures can be, and sometimes have been, proposed. I will not explore these questions here, but merely note that incompatibility of such proposals with the theory of phrase structure grammar stands as no barrier to them“ (Chomsky 1982, p.14f).[2]

Now, trees are a special type of graphs. Following Cormen et al. (1990, p.86ff), one may also call the objects defined in (5) “rooted, directed, anticyclic, ordered graphs

---

[2] This passage is also quoted by Blevins (1990, p.4), who employs graphs, called “mobiles,“ that allow both discontinuous constituents and MC/MD. A different type of multidimensional phrase-structure is employed by Haegeman&van Riemsdijk (1986) in their analysis of verb projection raising. Their mechanism of “reanalysis“ allows a sequence of categories to possess several (conflicting) phrase-structural analyses at one and the same time. See Kolb (1997b) for arguments that even the proper analysis of adjunction requires structures richer than trees.

(with self-loops).“[3] The set-based objects defined in Chomsky (1995a, chapter 4), and the modified objects I will introduce in section 3.3, could be translated into rooted, directed, acyclic, (unordered) graphs. Of course, it has to be shown how linear precedence can be computed on the basis of unordered graphs, a question that equally arises for the set-based minimalist structures (cf. sections 2.6.2 and 3.4.3).

The first step toward MC/MD, however, consists in abandoning the so-called “Single Mother Condition“ (SMC), (8), sometimes also called “Nonlooping Condition.“[4] From the perspective of trees, this condition is definable in terms of “immediate dominance“ (7), itself based on dominance via the notion of “proper dominance“ (6).

(6) Proper Dominance
$(\forall x,y \in N)[\ xPDy \leftrightarrow xDy \wedge x \neq y\ ]$

(7) Immediate Dominance
$(\forall x,y \in N)[\ xIDy \leftrightarrow (\ xPDy \wedge \neg\ (\exists z \in N)[\ xPDz \wedge zPDy\ ]\ )\ ]$

(8) Single Mother Condition (SMC)
$(\forall x,y,z \in N)[\ (\ xIDz \wedge yIDz\ ) \rightarrow x = y\ ]$

What keeping or abandoning (8) implies can be made more tangible if one considers the directedness of immediate dominance. Thus, in the language of graph-theory, a pair of nodes <x,y> (∈ ID) defines an “edge“ that “leaves“ x and “enters“ y. One can count the number of edges entering a node, arriving at the “in-degree“ of a node, and the number of edges leaving a node, the “out-degree“ (Cormen et al. 1990, p.87). Minimalist syntax takes the in-degree of nodes to lie invariably at 1, except for root nodes, whose in-degree is 0. The out-degree lies invariably at 2, i.e. structures are “binary branching,“ except for terminals, whose out-degree equals 0.[5] Thus, abandoning the Single Mother Condition amounts to allowing in-degrees of any finite integer n, ($n \geq 1$), except for root nodes, which by definition keep their in-degree at 0.

Formally, given (6) and (7), the SMC is a logical consequence of the definition of trees in (5).[6] To show this, we assume (9) holds.

(9) $aIDc \wedge bIDc \wedge a \neq b$

Applying the definitions of ID and PD, we can derive (10).

(10) $\neg(aDb) \wedge \neg(bDa)$

---

3 “Directedness“ is usually represented not by arrows but by top-down orientation of the graph on the page. “Anticyclicity“ seems to be an appropriate term for graphs that only allow cycles of length 1, i.e., cycles that are “self-loops.“

4 Cf. McCawley (1968, p.244), Barker&Pullum (1990, p.20), and Blevins (1990, p.48).

5 For a recent approach appealing to ternary branching see Brody (1998).

6 Cf. Barker&Pullum (1990, p.20).

Thus, nodes violating the SMC could not be ordered wrt to each other in terms of D. It then follows from the Exclusivity Condition on trees that they must be ordered wrt each other in terms of P, that is, (11) holds.

(11) (aPb) $\vee$ (bPa)

From the Nontangling Condition on trees we know that ordering two nodes in terms of P orders whatever these nodes dominate in the same way. Thus, given (9) and (11) we get (12).

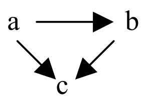
(12) cPc

(12) is in conflict with the requirement on P to be irreflexive. So we derive contradiction (13) and have an indirect proof that the SMC follows from the definitions of "constituent structure tree," PD, and ID.

(13) cPc $\wedge$ $\neg$(cPc)

As (10) indicates, one special case of MC/MD is filtered out by the definition of immediate dominance right away.[7] Thus, consider the following graph.

(14)



This could represent the proper dominance relation given in (15).

(15) aPDb $\wedge$ aPDc $\wedge$ bPDc

(15), however, directly prevents an interpretation of (14) as representing an ID relation, given the definition of immediate dominance. The (minimal) ID representation compatible with (15) is (16) instead.

(16) a → b → c

Crucially, (14) is a subcase of the kind of structure that canonically arises if movement constructions or chains are captured by means of MC/MD. This has been indicated in section 2.7.3 already. Thus, if one wants to interpret the relation in (14) as immediate dominance relation, it is not enough to abandon the SMC. ID has to be taken as a

7 Thanks to Tom Cornell (p.c.) for directing my attention to this issue.

defining primitive of the graphs employed. Let us call the appropriate structures “rooted MC/MD-graphs.“[8]

(17) A *rooted MC/MD-graph* is a pair <N,ID>, where
N is a finite set, the set of nodes,
ID is a binary relation on N, the immediate dominance relation,
and such that
(1) ID^+ is irreflexive (Acyclicity)
(2) (∃x ∈ N)(∀y ∈ N)[<x,y> ∈ ID*] (Single Root Condition)

ID^+ will also be called “proper dominance relation,“ PD, and ID* “dominance relation,“ D. On the basis of (17), (14) and (16) can now be distinguished insofar as they represent different rooted MC/MD-graphs, namely, the ones in (18a) and (18b) respectively.

(18) a. <{a, b, c}, {<a,b>, <a,c>, <b,c>}>
b. <{a, b, c}, {<a,b>, <b,c>}>

Clearly, this distinction could not be made on the basis of D (= ID*), where they “collapse“ into (19).[9]

(19) a. <{a, b, c}, {<a,a>, <a,b>, <a,c>, <b,b>, <b,c>, <c,c>}>

Let us now turn to the issue of precedence and its relation to MC/MD structures. Most radically, one could dispense with P, i.e. work with unordered graphs like the ones defined in (17) (cf. Barker&Pullum 1990). There would then be no equivalent to the Exclusivity Condition or Nontangling Condition on trees. What is required in this case is some kind of sorting algorithm that linearizes the terminals at the (interface to the) PHON-component. As already indicated in section 2.6.2, Chomsky (1995a, section 4.8) proposes to implement Kayne's (1994) LCA, for the purpose of translating asymmetric c-command into linear precedence.

Alternatively, one can define *ordered* MC/MD structures, two variants of which I will discuss here. The first one, developed by McCawley (1968) and employed by Blevins (1990), consists in giving a strict linear ordering of *terminal* nodes. On the basis of this ordering, a partial precedence relation on non-terminals can be induced. The result is constrained by a “Partial Exclusivity Condition.“ The second approach is taken by Peters&Ritchie (1981), who define a strict linear ordering for each set of *sister* nodes in their own version of MC/MD structures, called “linked trees.“

Let us start off with the former system. I'll develop this approach in terms of MC/MD-graphs as defined in (17). Thus, assume we have added a strict partial ordering

---

8 Cf. the definition of “acyclic graphs“ in Kracht (2001). I disregard labeling here. I adopt the convention that for any relation R, R^+ denotes the transitive closure of R and R^* denotes the reflexive transitive closure of R.

9 The definitions of “labeled graph“ (Barker&Pullum 1990, p.20) and “mobile“ (Blevins 1990, p.49) seem to neglect this point.

on N, the "precedence relation" P, and thereby obtained a triple <N,ID,P>, the basis of "*ordered*, rooted MC/MD-graphs."
We then define the set of terminals, T, as a subset of N.

(20) Terminals
$(\forall x \in N)[\ x \in T \leftrightarrow \neg(\exists y \in N)[\ \langle x,y\rangle \in ID\ ]\ ]$

Next, we require the set of terminals to be linearly ordered (cf. Blevins 1990, p.50).

(21) $(\forall x,y \in T)[\ x \neq y \rightarrow (\ \langle x,y\rangle \in P \vee \langle y,x\rangle \in P\ )\ ]$

A linguistically sound projection of P onto non-terminals further requires a weakened Exclusivity Condition, given in (22).

(22) Partial Exclusivity Condition (PEC)
$(\forall x,y \in N)[\ \langle x,y\rangle \in ID^* \rightarrow (\ \langle x,y\rangle \notin P \wedge \langle y,x\rangle \notin P\ )\ ]$

Stating the "induction" principle, i.e. the "Precedence Inheritance Condition" requires some additional preparation.

Let's call the set of nodes dominated by a node x the "lower cone" of x, abbreviated $|\!\!\downarrow x$ (Cf. Grefe&Kracht 1996, p.3).

(23) Lower Cone
$|\!\!\downarrow x =_{def} \{\ y \mid \langle x,y\rangle \in ID^*\ \}$

Since $|\!\!\downarrow x$ contains the nodes that make up the constituent with root node x, we will also speak of the "constituent" $|\!\!\downarrow x$.[10] Also, we need the collection of nodes *properly* dominated by x, i.e. the "proper lower cone" of x, abbreviated as $\downarrow x$.

(24) Proper Lower Cone
$\downarrow x =_{def} \{\ y \mid \langle x,y\rangle \in ID^+\ \}$

We further distinguish terminals from non-terminals wrt their "orderable lower cone," abbreviated as $\Downarrow x$.

(25) Orderable Lower Cone

$$\Downarrow x =_{def} \begin{cases} |\!\!\downarrow x, \text{ if } x \in T \\ \downarrow x, \text{ if } x \in N\text{-}T \end{cases}$$

[10] Technically, it may be advantageous to define constituents not just as sets of nodes, but as (sub-)trees or (sub-)graphs, as done in Grefe&Kracht (1996, p.3). For an updated version of the latter, see Kracht (1999).

We now generalize precedence for sets of nodes in the following way. (X and Y are set variables.)

(26) Set Precedence

$<X,Y> \in \mathbb{P} \Leftrightarrow_{def} (\forall x,y \in N)[\ (x \in X \wedge y \in Y) \rightarrow <x,y> \in P\ ]$

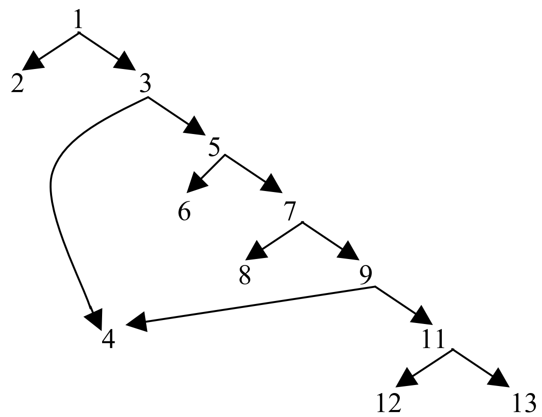
Finally, we can state the crucial "Precedence Inheritance Condition."[11]

(27) Precedence Inheritance Condition (PIC)

$(\forall x,y \in N)[\ <x,y> \in P \leftrightarrow <\Downarrow x,\Downarrow y> \in \mathbb{P}\ ]$

The consequence of the PIC is that precedence relations among terminals will be "inheritable" into regions of the MC/MD-graph that aren't "disturbed" by multidominance. Let's have a look at an example in order to see what that means.

(28)



(28) shows the kind of structuring, we are essentially concerned with, given the purpose of capturing (unbounded) dependencies. Let us assume that the order of terminals is <2,6,8,4,12,13>. In linguistic terms, this could be taken to mean that no "overt raising" of 4 has taken place. The complete relation P on the set of terminals is given in (29).

(29) $P_T = \{<2,6>, <2,8>, <2,4>, <2,12>, <2,13>, <6,8>, <6,4>, <6,12>, <6,13>, <8,4>, <8,12>, <8,13>, <4,12>, <4,13>, <12,13>\}$

This fulfills condition (21). Likewise, the PEC is (trivially) complied with, given that for each terminal x, $<x,x> \in ID^*$ and $<x,x> \notin P$, the latter preserving the irreflexivity of P. In addition, precedence inheritance is fulfilled, given the identity of $\Downarrow x$ and $\{x\}$ in

[11] The PIC is closely related to the Nontangling Condition on trees. This is revealed if the latter is reformulated as follows.

(i) Nontangling Condition (reformulated)

$(\forall x,y \in N)[\ <x,y> \in P \rightarrow <\Downarrow x,\Downarrow y> \in \mathbb{P}\ ]$

the case of terminals. For example, the PIC states that <2,6> ∈ P iff <{2},{6}> ∈ ℙ. The latter has been defined to be equivalent to the former via “Set Precedence.“

Let's then have a look at the nonterminal 11. Can it be ordered wrt 4 for example, i.e. can we get <4,11> ∈ P, or <11,4> ∈ P? Clearly, we have <{4},{12,13}> ∈ ℙ, which is equivalent to <⇓4,⇓11> ∈ ℙ. Thus, we get <4,11> ∈ P “by induction.“

Exactly the same reasoning holds wrt 11 and the other terminals preceding 4. We can thus add <2,11>, <6,11>, and <8,11>. Next, if <8,11> ∈ P holds, <⇓8,⇓9> ∈ ℙ also holds, given that <{8},{4,11,12,13}> ∈ ℙ. The remainder of P will be completed in the same way.

Thus far, multidominance hasn't had any effect on the construction of P. Things change, however, if we reorder terminal 4 in (28) wrt the other terminals, such that ordering <2,4,6,8,12,13> results.

(30) P_T = {<2,4>, <2,6>, <2,8>, <2,12>, <2,13>, <4,6>, <4,8>, <4,12>, <4,13>, <6,8>, <6,12>, <6,13>, <8,12>, <8,13>, <12,13>}

Inducing precedence for 11 will work as before. However, integrating 9 is not straightforward. In order to get <8,9> ∈ P we need <{8},{4,11,12,13}> ∈ ℙ. This is impossible, since <4,8> ∈ P, not <8,4>. Crucially, 4 ∈ ⇓9 still holds, although 4 has somehow “left“ the constituent with root 9 as far as computing linear precedence is concerned. Conversely, <9,8> ∈ P cannot hold either, given that among other things 12 ∈ ⇓9 but <8,12> ∈ P. Thus, 8 and 9 cannot be ordered wrt each other in terms of P. Indeed, 8 and 9 cannot be ordered wrt each other at all under these conditions, since <8,9> ∉ ID* and <9,8> ∉ ID*.

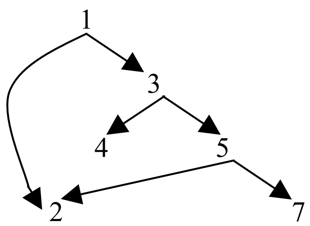
We now see why the Exclusivity Condition on trees has to be replaced by the Partial Exclusivity Condition if trees are replaced by ordered rooted MC/MD-graphs. Nodes that aren't ordered wrt each other in terms of ID* may be not ordered wrt each other in terms of P under specific circumstances.

Before going into the description of these “specific circumstances,“ let me repeat the main point of what has been established. One way of dealing with linear precedence within MC/MD syntax is to define *ordered* rooted MC/MD-graphs on top of their unordered counterparts. The result looks as follows (cf. McCawley 1968, Blevins 1990).

(31) An *ordered rooted MC/MD-graph* is a triple $\langle N,ID,P \rangle$, where
N is a finite set, the set of nodes,
ID is a binary relation on N, the immediate dominance relation,
P is a strict partial ordering on N, the precedence relation,
and such that
(1) ID^+ is irreflexive (Acyclicity)
(2) $(\exists x \in N)(\forall y \in N)[\ \langle x,y \rangle \in ID^*\ ]$ (Single Root Condition)
(3) $(\forall x,y \in T)[\ x \neq y \rightarrow (\ \langle x,y \rangle \in P \vee \langle y,x \rangle \in P\ )\ ]$
(Obligatory Ordering of Terminals)
(4) $(\forall x,y \in N)[\ \langle x,y \rangle \in ID^* \rightarrow (\langle x,y \rangle \notin P \wedge \langle y,x \rangle \notin P\ )\ ]$
(Partial Exclusivity Condition)
(5) $(\forall x,y \in N)[\ \langle x,y \rangle \in P \leftrightarrow \langle \Downarrow x, \Downarrow y \rangle \in \mathbb{P}\ ]$
(Precedence Inheritance Condition)

Now, multidominance results in P-unorderable pairs of nodes whenever there is "discontinuity." Take a very simple case like (32).

(32)



If the order of terminals in (32) is $\langle 2,4,7 \rangle$, $\downarrow$5 is a discontinuous constituent. The yield of $\downarrow$5, i.e. 2+7, is not a substring of the entire string. For the yield of $\downarrow$1 is 2+4+7. 4 interrupts 2+7. We have already seen that it follows within ordered rooted MC/MD-graphs as defined in (31) that 4 and 5 are P-unorderable. $\langle \Downarrow 4, \Downarrow 5 \rangle \notin \mathbb{P}$ and $\langle \Downarrow 5, \Downarrow 4 \rangle \notin \mathbb{P}$. The general principle underlying "P-unorderability" for pairs of nodes not ordered wrt each other in terms of ID* is given in (33).[12] Call this the "Precedence Disinheritance Condition."[13]

---

[12] Recall that the ones that are ordered in terms of ID* are P-unorderable because of the Partial Exclusivity Condition.

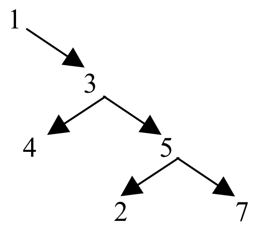
[13] The PDC follows from the PIC. Thus assume (i).
(i) $\langle a,b \rangle \in P \wedge \langle b,c \rangle \in P \wedge \langle d,a \rangle \in ID^* \wedge \langle d,c \rangle \in ID^* \wedge \langle d,b \rangle \notin ID^* \wedge$
$(\langle b,d \rangle \in P \vee \langle d,b \rangle \in P)$
Take (ii) as our first case. (ii) $\langle b,d \rangle \in P$. From this we derive (iii), given the PIC. (iii) $\langle \Downarrow b, \Downarrow d \rangle \in \mathbb{P}$. From (i) we can also derive (iv), given the PIC. (iv) $\langle \Downarrow a, \Downarrow b \rangle \in \mathbb{P}$. Now there are four subcases to consider: a) $a \in T$ and $b \in T$, b) $a \in T$ and $b \notin T$, c) $a \notin T$ and $b \in T$, and d) $a \notin T$ and $b \notin T$. Starting off with a), we derive (v) from (i). (v) $a \in \Downarrow d$. From (v) and (iii) together we get (vi), which contradicts (i). (vi) $\langle b,a \rangle \in P$ [$\bot$]. Considering subcase b), assume that $b^* \in \downarrow b$. From (iii) and (v) we derive (vii). (vii) $\langle b^*,a \rangle \in P$. However, (iv) yields (viii), which contradicts (vii). (viii) $\langle a,b^* \rangle \in P$ [$\bot$].

(33) Precedence Disinheritance Condition (PDC)

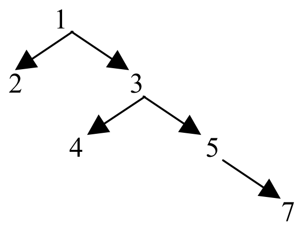
$$(\forall w,x,y,z \in N)[\ (\ \langle x,y\rangle \in P \wedge \langle y,z\rangle \in P \wedge \{x,z\} \subseteq \downarrow w \wedge y \notin \downarrow w\ ) \rightarrow (\ \langle y,w\rangle \notin P \wedge \langle w,y\rangle \notin P\ )\ ]$$

Before introducing the second approach to precedence within MC/MD syntax, let us briefly look at the treatment of unbounded dependencies in terms of standard "movement." Thus, consider (34).

(34) a.



b.



Construed in the most radical way, movement displaces a constituent. The essential difference between (32) and (34) can be captured in terms of the "proper upper cone" of a node.

(35) Proper Upper Cone

$$\uparrow x =_{def} \{y \mid \langle y,x\rangle \in ID^{+}\}$$

The transition from (34a) to (34b) implies that $\uparrow 2$ is altered, as stated in (36).

(36) $\uparrow 2 = \{5,3,1\} \quad \Rightarrow \quad \uparrow' 2 = \{1\}$

Now assume that instead of deriving (34b) from (34a), (32) is derived. Then, $\uparrow 2$ remains constant.

(37) $\uparrow 2 = \{5,3,1\} \quad \Rightarrow \quad \uparrow' 2 = \{5,3,1\}$

Turning to case c) assume that $a^* \in \downarrow a$. (iii) and (v) then yield (ix). (ix) $\langle b,a^*\rangle \in P$. From (iv) we derive (x), which contradicts (ix). (x) $\langle a^*,b\rangle \in P$ [$\perp$]. Finally, in order to deal with case d) we assume that $a^* \in \downarrow a$ and $b^* \in \downarrow b$. From (iii) and (v) we conclude that (xi) holds. (xi) $\langle b^*,a^*\rangle \in P$. (iv), however, yields (xii), which contradicts (xi). (xii) $\langle a^*,b^*\rangle \in P$ [$\perp$].

Thus, we can conclude that assuming (ii) is incompatible with the PIC. In order to complete the indirect proof that the PDC follows from the PIC, we have to show that (xiii) as part of (i) is equally incompatible with the PIC. (xiii) $\langle d,b\rangle \in P$ This can be shown in exactly parallel fashion, provided that c is replaced for a.

Drawing on the notion of "proper upper cone" (see (35), below) the effect captured by the PDC can be formulated somewhat more elegantly, but likewise more redundantly, in terms of the following principle. (R is a variable over P, "precedence," and its complement P', "succession.")

(xiv) Principle of Orientation Disturbance

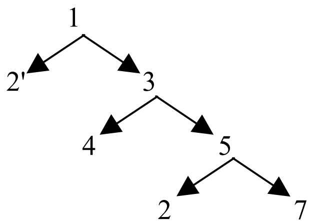
$$(\forall x,y,z \in N)[\ (\ \langle x,y\rangle \in R \wedge z \in \uparrow x\ ) \rightarrow \langle y,z\rangle \notin R\ ]$$

The type of dependencies just discussed observes the widely assumed c-command condition, which could therefore be rephrased as follows.[14]

(38) C-Command Condition on Dependencies
a. Move: $\uparrow x \subset \uparrow' x$
b. Multidominance: $\uparrow x = \uparrow' x$

However, as amply discussed earlier, minimalist syntax takes movement to involve copying and a chain relation between the “voyager“ and its copy left behind. Thus, the “real“ transition from (34a) under such a construal produces (39).

(39)



Again, c-command is observed, i.e. $\uparrow 2 \subset \uparrow 2'$. We have also already seen that identity is not the proper relation between 2 and 2' (cf. section 2.7.1). An identification of 2 and 2' is what underlies (32) instead, i.e. trees would have to be replaced by MC/MD-graphs.[15]

The c-command condition on dependencies expressed in terms of multidominance has been discussed in Barker&Pullum (1990, p.22) under the name of “Connected Ancestor Condition.“ I adapt their formulation to MC/MD-graphs.[16]

---

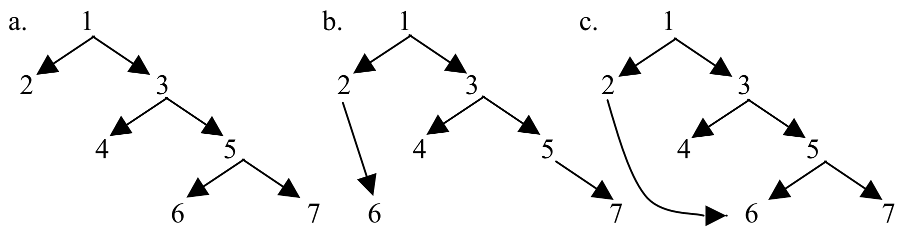
[14] Monotonicity, constancy, or “syntactic compositionality“ (Peter Staudacher p.c.), as revealed by this formulation of c-command, are all potential sources of simplicity. See Chametzky (1996) and Epstein (1999) for attempts to “explain“ these properties of c-command.

[15] See section 3.3 for further discussion of this. A formal analysis of the chain relation is beyond the scope of this work. See Kracht (2001). Let me simply note that one way of looking at chains would be to add another dimension to trees, e.g. a reflexive, transitive, symmetric relation E on N, called “enchain.“ Reflexivity gives us trivial chains per default. In addition one has to stipulate that copying leads to “enchainment.“ This could be formulated as follows.
(i) $(\forall x,y \in N)[\ Copy(x,y) \rightarrow <x,y> \in E\ ]$
This gives rise to the following relation E for the tree in (39).
(ii) $E = \{<1,1>, <2',2'>, <3,3>, <4,4>, <5,5>, <2,2>, <7,7>, <2',2>, <2,2'>\}$
Of course, it is desirable to define chains over constituents rather than nodes. Also the copy relation is in need of further clarification. See Kracht (2001). For linguistic purposes, chains involving traces, resumptive pronouns, or parasitic gaps, as well as binding chains would have to be reviewed.

[16] See Barker&Pullum (1990) for the equivalence of MC/MD-graphs observing the CAC to trees wrt formal command properties. Under the obvious definition of “Upper Cone“ in (i), the CAC can be reformulated as in (ii).
(i) Upper Cone
$\Uparrow x =_{def} \{y \mid <y,x> \in ID^*\}$

(40) Connected Ancestor Condition (CAC)
$(\forall x,y,z \in N)[\ (\langle x,z\rangle \in ID \wedge \langle y,z\rangle \in ID) \rightarrow (\langle x,y\rangle \in ID^* \vee \langle y,x\rangle \in ID^*)\ ]$

Let us finally have a look at structures deviating from (38). Thus, transitions from (41a) to (41b)/(41c) can be defined.

(41)


(42) a. $\uparrow 6 = \{5,3,1\} \quad \Rightarrow \quad \uparrow' 6 = \{2,1\}$ [= (41b)]
b. $\uparrow 6 = \{5,3,1\} \quad \Rightarrow \quad \uparrow' 6 = \{5,3,2,1\}$ [= (41c)]

This time, $\uparrow' x$ includes "new information." Under displacement this leads to incongruent proper upper cones, i.e. $\uparrow x \not\subset \uparrow' x$ and $\uparrow' x \not\subset \uparrow x$ and $\uparrow x \neq \uparrow' x$. Again, under multidominance no information is lost. Thus $\uparrow x \subset \uparrow' x$.

The additional complication arising in (41) is, of course, correlated with the fact that from a derivational perspective, these transitions are counter-cyclic (cf. section 2.5.2). Only the former preserve c-command or the CAC. Thus, given my hypothesis 3, transitions like (41) will have to be avoided. The one problematic case, standing in the way of such an assumption, is X°-movement, to which I will return below (cf. 3.4.2).[17]

An alternative approach to linear order within MC/MD-syntax is developed in "phrase linking grammar" (PLG) as defined by Peters&Ritchie (1981) and adopted by Engdahl (1986).[18] The main innovation of PLG is to replace trees by "linked trees," which are defined as follows (Peters&Ritchie 1981, p.6; Engdahl 1986, p.44f).

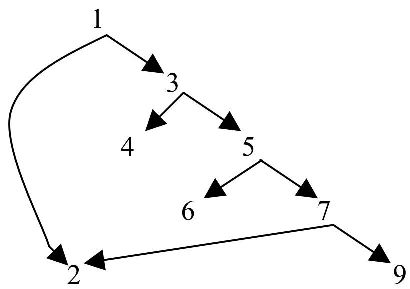
(ii) Connected Ancestor Condition [reformulated]
$(\forall x,y,z \in N)[\ (x \in \uparrow z \wedge y \in \uparrow z) \rightarrow (x \in \Uparrow y \vee y \in \Uparrow x)\ ]$

[17] Another logically possible transition can be formulated as in (i).
(i) $\uparrow x \subseteq \uparrow' x$
This type of relationship is commonly called "lowering." An example would be the transition from (34b) to (34a), i.e. a reversal of the standard operation. Looking at it in terms of multidominance is one way toward trivializing the directionality involved in lowering. See Haider (1997) for discussion of whether the need to rule out such artifacts counts against the dynamical perspective of derivations.

[18] See also Joshi (1985) for some discussion of the formal properties of PLG and of how to add links to his own TAG framework.

(43) A *linked tree* is a finite set *N* of nodes (vertices) together with binary relations *I* (of immediate tree domination) and *L* (of immediate link domination) on *N*, and functions *P* (of left-to-right precedence) and *f* (which labels nodes with vocabulary symbols) having domain *N* and ranges contained respectively in $N \times N$ and $V_T \cup V_N$ satisfying conditions (i) - (v):
(i Linear Precedence Ordering of Siblings) P(n) is a strict linear ordering of {m| <n,m> ∈ I∪L} for all n in N,
(ii Root) there is an r in N such that <r,n> ∈ I* for all n ∈ N,
(iii Unique Tree Parent) I^{-1} is a partial function defined just at members of N - {r},
(iv Tree Parent Dominates Link Parent(s)) if <n,n'> ∈ L, then there are m_0, . . . , m_p ∈ N (p>0) such that $m_1 \neq$ n', m_p = n, <m_0,n'> ∈ I, and <m_i,m_{i+1}> ∈ I whenever 0 ≤ i < p, for all n, n' ∈ N.
(v Node Labeling) *f*(n) ∈ V_N iff there is an n' in N such that <n,n'> ∈ I∪L for all n in N.

Now, instead of treating all instances of immediate dominance alike, as is the case with ID within MC/MD-graphs, linked trees assign a unique designated “tree parent“, I^{-1}(n), to each node *n*, except for *r*. Thus, the surface- or Spell-Out position of *n* can be determined on the basis of this tree parent. In addition, traces are replaced by “links“ in the sense that for each *m* that would have dominated a trace of *n* in a tree, <m,n> ∈ L in a linked tree.
Let us have a look at the following simple example.

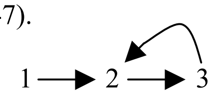
(44)


Instead of a single ID relation, PLG takes (44) to be partitioned into I and L as follows.

(45) a. I = {<1,2>, <1,3>, <3,4>, <3,5>, <5,6>, <5,7>, <7,9>}
b. L = {<7,2>}

Now, condition (iv) imposes a restriction on linked trees, similar to the Connected Ancestor Condition (CAC). Thus, take n=7 and n'=2 in (44)/(45), i.e. <7,2> ∈ L. The tree parent of 2 is 1, i.e. m_0=1 and <1,2> ∈ I. Then there is a sequence of nodes from the tree parent m_0 (=1) to the link parent m_p (=n=7), where each successor is related to its predecessor in terms of I, i.e. <m_i,m_{i+1}> ∈ I. Such a sequence is minimally of length

2, i.e. $p>0$. Thus, tree parent and link parent cannot be identical.[19] The constraint that $m_1 \neq n'$, i.e. that the I-path from tree parent to link parent must not go via their common child, together with condition (iii) guarantees the acyclicity of the resulting graph.[20] In (44) this is redundant. Assume $m_1=2$. Then $\langle 2,7\rangle \in I$ must hold, which violates the uniqueness condition (iii), given that $\langle 5,7\rangle \in I$ too. However, consider the graph in (46)/(47).

(46) 

(47) a. $I = \{\langle 1,2\rangle, \langle 2,3\rangle\}$
b. $L = \{\langle 3,2\rangle\}$

---

[19] Technically, this means that linked trees are not “multigraphs,“ that is graphs allowing for the possibility of “several edges joining the same pair of vertices“ (Trudeau 1993, p.24).

[20] A proof of acyclicity for linked trees would have to show that a) I^+ is irreflexive and b) adding links preserves irreflexivity, i.e. $(I \cup L)^+$ is irreflexive too (cf. the definition of MC/MD-graphs). The latter property may actually not hold, as shown in the following footnote. Here I give the outlines of a proof for a). Assume the definitions of proper lower and upper cone are transposed to I, i.e. we can use the same notation as before. We then have to establish (i).

(i) $\neg(\exists x \in N)[\ x \in {\downarrow}x\ ]$

Of course, (i) could be formulated differently, given the following equivalence.

(ii) $(\forall x \in N)[\ x \in {\downarrow}x \leftrightarrow x \in {\uparrow}x\ ]$

Let us now characterize a “loop“ as a set of nodes in a graph where each node is a member of its own proper lower (and upper) cone and all members have identical lower and upper cones. This is what underlies the notion of “loop set“ of x.

(iii) Loop Set

$\infty x =_{def} \{y \mid y \in {\downarrow}y \wedge {\downarrow}x = {\downarrow}y \wedge {\uparrow}x = {\uparrow}y\}$

It can be shown that a non-empty loop set of x must contain x itself, i.e. (iv) holds.

(iv) $(\forall x \in N)[\ \infty x \neq \varnothing \rightarrow x \in \infty x\ ]$

Likewise, if I is giving rise to a loop set ∞x, then every member of ∞x must be connected to at least one member of ∞x via I, i.e. (v) holds.

(v) $(\forall x,y \in N)[\ y \in \infty x \rightarrow (\exists z \in N)(\ z \in \infty x \wedge \langle z,y\rangle \in I\ )\ ]$

Showing (i) then boils down to showing (vi), i.e. there are no loops.

(vi) $(\forall x \in N)[\ \infty x = \varnothing\ ]$

To show this one has to consider two cases: a) non-empty loops containing r, and b) non-empty loops excluding r. The former case is directly ruled out by the fact that if $r \in \infty x$ then $r \in {\downarrow}r$. This violates condition (iii) of (43), which takes $I^{-1}(r)$ to be undefined. Case b) is excluded because such a loop has to be connected to the root at some point, given condition (ii) of (43). However, the node which establishes contact to the root must violate uniqueness condition (iii) of (43), since it will be immediately dominated by a node outside of ∞x as well as by another one inside of ∞x, the latter due to (v). The state of affairs conflicting with condition (iii) of (43) is formally expressed in (vii).

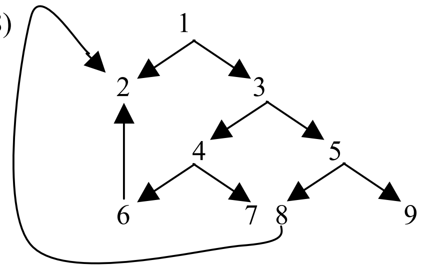
(vii) $(\forall x \in N)[\ (\ \infty x \neq \varnothing \wedge r \notin \infty x\ ) \rightarrow$
$(\exists w,y,z \in N)(\ w \in {\uparrow}x - {\downarrow}x \wedge \{y,z\} \subseteq \infty x \wedge \langle w,y\rangle \in I \wedge \langle z,y\rangle \in I\ )\ ]$

Here 3 does not have a second tree parent. This kind of cycle is directly ruled out by the condition that $m_1 \neq n'$, i.e. there must be a path from tree parent to link parent which does not go via the link child.[21]

Importantly, condition (iv) differs from the CAC in two main respects. First, it is assumed that the designated ("principal") tree position of each node is its "highest" position. If designated position and surface position are identified, this assumption has to be modified for minimalist syntax, where abstract (post Spell-Out) dependencies abound.[22]

Secondly, while each link parent must be dominated by the tree parent, link parents need not be ordered wrt each other in terms of I*. This makes PLG capable of capturing structures where one constituent seems to license two or more gaps, as is often assumed for ATB and parasitic gap constructions.[23] Thus consider (48)/(49).

(48)



(49) a. I = {<1,2>, <1,3>, <3,4>, <3,5>, <4,6>, <4,7>, <5,8>, <5,9>}
b. L = {<6,2>, <8,2>}

Condition (iv) is satisfied for both elements of L in (49b). Nothing more is required. Thus, neither <6,8> $\in$ I* nor <8,6> $\in$ I*. The CAC, on the other hand, is violated by the graph in (48)/(49).[24]

---

[21] By the same mechanism, all trivial links of type <x,x> where x $\neq$ r are ruled out. Finally, trying to make r a link child is prohibited by the fact that there would be no tree parent for the link child. This is in conflict with condition (iv) again. It seems, however, that cycles may arise if two (or more) links are added. Thus consider the following linked tree.
(i) a. I = {<1,2>,<1,3>,<2,4>,<2,5>,<3,6>,<3,7>} b. L = {<5,3>,<6,2>}
Node 3 is link child of a child of its sister 2, and, conversely, 2 is link child of a child of 3. The tree parent for both links is 1. The path from tree parent to link parent is <1,2,5> for link <5,3> and <1,3,6> for link <6,2>. Condition (iv) is fulfilled. Yet, a cycle has been created, comprising the set {2,5,3,6}. As a result $(I \cup L)^+$ of linked trees is not irreflexive, as opposed to ID^+ of (ordered) rooted MC/MD-graphs.

[22] It would be an interesting formal property of minimalist syntax if abstract movements were invariably carried out in a single step, i.e. if there were no successive-cyclic LF movement. Then the surface position of each constituent could be determined on the basis of λ by a very simple algorithm.

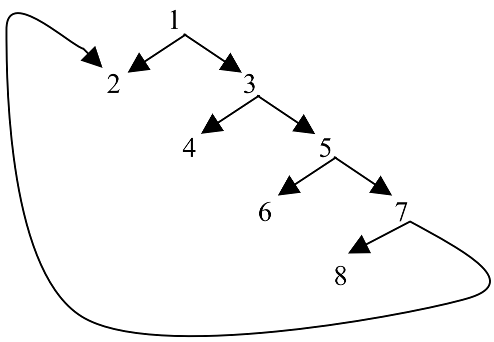
[23] For the latter see Engdahl (1986, p.129ff).

[24] It would have to be shown whether Barker&Pullum's (1990) results concerning formal command properties of acyclic graphs obeying the CAC carry over to linked trees.

Since the additional complexity of such multiple dependencies is beyond the immediate concerns of (narrow) minimalist syntax, the CAC will later be assumed to hold for the MC/MD-structures I introduce to simulate minimalist syntax.[25]

On the other hand, both condition (iv) and the CAC rule out dependencies that violate strictest c-command. We have seen that for the CAC above. Now assume we add link <8,6> to L of (48)/(49). There is no "descending" path from 4, the tree parent of 6, to 8, the link parent of 6, i.e. $<4,8> \notin I^*$. This violates condition (iv). Clearly, dependencies have to obey c-command within PLG. Thus, X°-movement as standardly construed cannot directly be captured in terms of linked trees either.

Let us now return to the issue of linear precedence. The approach based on the "Partial Exclusivity Condition" and the "Precedence Inheritance Condition" discussed above requires a strict ordering of *terminals*. This ordering can be projected onto non-terminals dominating continuous constituents. The PLG approach starts off with an ordering of each set of *siblings*, as stated in condition (i) on linked trees. Let us call the generalized union of these orderings "sibling precedence" (SP). Now consider the graph in (50), whose SP relation is given in (51).

(50)


(51) SP = {<2,3>, <4,5>, <6,7>, <8,2>}

Assume for convenience that the definitions of cones are transposed from ID to I and $I \cup L$, so that the same notation can be used in the following. Clearly, the equivalent of the Partial Exclusivity Condition on MC/MD-graphs does not hold for linked trees. $<2,3> \in$ SP and $<3,2> \in (I \cup L)^*$. Likewise the order of terminals is not unambiguously determined by (50)/(51). A simple recursive projection of (general) precedence like

[25] Nunes (1995, 2001) analyzes parasitic gaps (and ATB constructions) in terms of "sideward movement." Crucially, as one of the derivable consequences of his system, the two gaps are not part of the same chain. There must be separate chains for each gap. Thus, the "anti-c-command" condition on parasitic gaps falls out as a consequence. While PLG requires separate "links" for distinct gaps as well, the anti-c-command condition would have to be stipulated, given that successive-cyclic movement will likewise be captured in terms of separate links. This time, however, one link parent will dominate the other. For further discussion of Nunes (1995), see Gärtner (1998).

(52), based on the Nontangling Condition on trees, is unsatisfactory because it does not (necessarily) result in a strict partial order.

(52) Precedence
a. $(\forall x,y \in N)[\ \langle x,y\rangle \in SP \rightarrow \langle x,y\rangle \in P\ ]$
b. $(\forall x,y,z,w \in N)[\ (\ \langle x,y\rangle \in P \wedge z \in \downarrow x \wedge w \in \downarrow y\ ) \rightarrow \langle z,w\rangle \in P\ ]$

For example, in (50)/(51) irreflexivity would be violated because $\langle 2,2\rangle \in P$ and asymmetry would be violated because both $\langle 2,4\rangle \in P$ and $\langle 4,2\rangle \in P$. However, PLG has a fairly simple way out.

> "The *language generated* by a phrase-linking grammar G is the set of yields (terminal strings) of trees that result by deleting all links from a linked tree admitted by G" (Peters&Ritchie 1981, p.9).

Thus, for the purpose of linearization one can define a restricted ordering condition, which takes into account only "tree siblings."

(53) Linear Precedence Ordering of Siblings
P(n) is a strict linear ordering of $\{m|\ \langle n,m\rangle \in I\}$ for all n in N

The resulting generalized union called "tree sibling precedence" (TSP) will be a subset of SP. Thus, since $\langle 7,2\rangle \in L$ of (50), TSP of (50) will be (54), i.e. $\langle 8,2\rangle \notin TSP$.

(54) $TSP = \{\langle 2,3\rangle, \langle 4,5\rangle, \langle 6,7\rangle\}$

Now (tree) precedence (TP) can be recursively projected as stated in (55), except that SP has to be replaced by TSP.[26]

---

[26] Here lower cones are taken to be interpreted on the basis of just I. The actual PLG approach to "precedence projection" is more complicated.

(i) A node m of a linked tree $\langle N,I,L,P,f\rangle$ is a *left-most descendent* of a node n in N if there are nodes $n'_1, \ldots, n'_p$ ($p\geq 1$) such that $n = n'_1$, $n'_p = m$, $\langle n'_i, n'_{i+1}\rangle \in I\cup L$ for $1\leq i<p$, and for no $i=1, \ldots, p-1$ is there an n'' such that $\langle n'',n_{i+1}\rangle \in P(n_i)$.

The concept of *right-most descendent* is defined in an analogous way (cf. Peters&Ritchie 1981, p.7). On the basis of (i) the left and right edges of each constituent are captured. Then it is possible to define a notion of adjacency, which is called "immediate precedence."

(ii) A node n of $\langle N,I,L,P,f\rangle$ *immediately precedes* a node m in N if there are n', m' and n_0 in N such that $\langle n_0,n'\rangle$ and $\langle n_0,m'\rangle$ are in $I\cup L$, n is a right-most descendent of n' and m is a left-most descendent of m', $\langle n',m'\rangle$ is in $P(n_0)$, and there is no n_1 such that $\langle n',n_1\rangle$ and $\langle n_1,m'\rangle$ are in $P(n_0)$ (Peters&Ritchie 1981, p.7f).

Although this adds further ordering for each graph, it does not for example establish any relation between nodes 4 and 6 in graph (50). Thus, while 6 is a left-most descendent of 5, 4 is not a right-most descendent of anything, given the irreflexivity of that notion. Thus although 4 and 6 would be candidates for "adjacency," this does not follow from (i) and (ii). In that sense the PLG system is still incomplete. A variant of this approach, originating with Blackburn&Meyer-Viol (1996), is

(55) Tree Precedence
a. $(\forall x,y \in N)[<x,y> \in TSP \rightarrow <x,y> \in TP]$
b. $(\forall x,y,z,w \in N)[(<x,y> \in TP \wedge z \in \downarrow x \wedge w \in \downarrow y) \rightarrow <z,w> \in TP]$

The resulting TP relation for the graph in (50) will be (56).

(56) $TP = \{<2,3>, <2,4>, <2,5>, <2,6>, <2,7>, <2,8>, <4,5>, <4,6>, <4,7>, <4,8>, <6,7>, <6,8>\}$

This unambiguously encodes the intended yield for (56), namely, 2+4+6+8. The PLG approach to linear precedence can thus be summarized by saying that the links of linked trees are just λ-relevant and not π-relevant.

Some such split is actually required in any system capturing (unbounded) dependencies in terms of multiple positions in graphs. Thus, standard movement approaches decide on π-relevance by realizing all but one position in a chain as phonologically null elements called "traces." If copies are employed instead, as is the case in minimalist sytax, as we have seen, there has to be a mechanism keeping track of the π-relevant copy within complex structures.[27]

I will come back to linear order in section 3.4.3 in the context of my own system, which coincides to a considerable degree with the PLG approach.

Further discussion of "arboreal" theories of MC/MD can be found in Blevins (1990), among them his own "Mobile Grammar," which we already mentioned. Although more emphasis is put on discontinuous constituents there, which are allowed in mobiles as well, MC/MD is empirically defended on the basis of data reminiscent of English *tough*-constructions (cf. section 2.5) found in Niuean, a language spoken in New Zealand.[28]

---

proposed in Kracht (2001). There the ID relation for binary-branching graphs is partitioned into ID_1 and ID_2, such that $ID = ID_1 \cup ID_2$, where $<x,y> \in ID_1$ iff y is the *left* daughter of x and $<x,y> \in ID_2$ iff y is the *right* daughter of x. It is clear how to translate the earlier notion of "immediate precedence" into these terms.

(iii) A node n of $<N, ID_1,ID_2>$ *immediately precedes* a node m in N if there are n', m' and n_0 in N such that $<n_0,n'> \in ID_1$, $<n_0,m'> \in ID_2$, $<n',n> \in ID_2^*$, and $<m',m> \in ID_1^*$.

Note that for every x, $<x,x> ID_1^*$ and $<x,x> ID_2^*$. On the basis of (iii) we can now say that in graph (50), interpreted as a binary MC/MD-graph in the sense of Kracht (2001), 4 "immediately precedes" 6.

27 For proposals see Nunes (1995), Groat&O'Neil (1996), and Uriagereka (1997,1999).

28 Blevins' radical conclusions wrt the nature of grammatical description raise issues of the kind indicated in section 1.1 above, which I prefer to sidestep. Consider the following quote: "Proposals for generating nonstandard representations are outlined at various points in this discussion, though it is assumed throughout that representational issues can be productively investigated independently of generation strategies. This assumption clearly conflicts with the standard generative practice of evaluating syntactic analyses in close conjunction with proposals for generating them. Nevertheless, if the central claims of this work are in the main correct, this would suggest that the generative emphasis on systems of rules and principles has substantially hindered rather than advanced the understanding of representational issues" (Blevins 1990, p.360).

3.2.1 Structure Sharing (An Excursus)

Let me now turn to another variety of MC/MD, namely, the one appealed to in "attribute-value" based theories of grammar, often referred to as "value sharing," "structure sharing," or "reentrancy." This terminology points to a function-theoretic or graph-theoretic background, as will become clear momentarily. The common ground of these theories is their all-encompassing use of feature structures, associated with, or even directly incorporating, standard constituent structures. Informally speaking, feature structures pair up features and values, where values themselves can be (sets of) feature-value pairs. Thus the feature *agreement* in (57) is assigned a complex value consisting of the features *number* and *person*, the former having value *singular*, the latter *third*.

(57)
$$\left[\begin{array}{l} \text{agreement} = \left[\begin{array}{l} \text{number = singular} \\ \\ \text{person = third} \end{array}\right]_{\text{b}} \\ \\ \end{array}\right]_{\text{a}}$$

A concise definition is given by Johnson (1988, p.18) under the name of "attribute-value structures."[29]

(58) An *Attribute-Value Structure* A is a triple $A = \langle F,C,\delta \rangle$, where F is a set, C is a subset of F, and δ is a partial function from F×F into F such that δ(c,f) is undefined for all c ∈ C and f ∈ F. The set F is called the set of *Attribute-Value Elements* of A, and the set C is called the set of *Constant Elements* of A. The class of attribute-value structures is called AVS.

On this view, the representation in (57), called an "attribute-value matrix" (AVM), corresponds to the attribute-value structure in (59).

(59) a. F = {agreement, number, person, singular, third, a, b}
b. C = F-{a,b}
c. δ(a,agreement) = b, δ(b,number) = singular, δ(b,person) = third

Note the use of arbitrary names for complex attribute-value elements. Clearly, the appeal to *δ* as a function opens up the possibility of "value sharing." Thus, imagine we add a complex element *c* to the structure in (57). Nothing then prohibits the addition of *δ(c,agreement)=b* to (57c), which amounts to saying that *a* and *c* coincide wrt the value of their agreement attributes. The corresponding AVM captures value-sharing in terms of coindexation, as shown in (60).[30]

---

[29] Johnson abstracts away from particular linguistic frameworks. As already indicated earlier, an introduction into GPSG, HPSG, and LFG can be found in Abeillé (1993). For detailed exposition and analyses see Gazdar et al. (1985), Pollard&Sag (1987, 1994), and Bresnan (ed.) (1982).

[30] See Shieber (1986) for discussion of various notations.

(60)
$$\begin{bmatrix} p = \begin{bmatrix} \text{agreement} = \textcircled{1}\begin{bmatrix} \text{number} = \text{singular} \\ \text{person} = \text{third} \end{bmatrix}_b \end{bmatrix}_a \\ q = \begin{bmatrix} \text{agreement} = \textcircled{1} \end{bmatrix}_c \end{bmatrix}_e$$

(60) involves a case of "token identity." There is exactly one *3rd person singular* entity. For linguistic purposes it is important to also allow for "type identical" "occurrences" of linguistic elements , i.e. elements that happen to coincide in their featural make-up (cf. section 2.6.1). Consider (61).

(61)
$$\begin{bmatrix} p = \begin{bmatrix} \text{agreement} = \begin{bmatrix} \text{number} = \text{singular} \\ \text{person} = \text{third} \end{bmatrix}_b \end{bmatrix}_a \\ q = \begin{bmatrix} \text{agreement} = \begin{bmatrix} \text{number} = \text{singular} \\ \text{person} = \text{third} \end{bmatrix}_d \end{bmatrix}_c \end{bmatrix}_e$$

(60) differs from (61) in that the values of $\delta(a,agreement)$ and $\delta(c,agreement)$ are *type*-identical in the latter and *token*-identical in the former. The use of two arbitrarily named attribute-value elements as opposed to coindexation guarantees that difference. Suppose, however, that complex elements were just sets of attribute-value pairs. Then, given extensionality of set theory, *b* and *d* would be identical, as shown in (62) (cf. Carpenter 1992, p.125).

(62) a. b = {<number,singular>,<person,third>}
b. d = {<number,singular>,<person,third>}

As a result, the structures in (60) and (61) would become identical, representable by the set in (63).

(63) {<p,{<agreement,{<number,singular>, <person,third>}>}>,
<q,{<agreement,{<number,singular>, <person,third>}>}>}

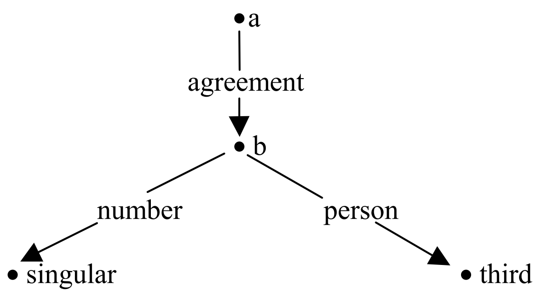
That the function-theoretic approach may not be the most perspicuous way of dealing with type identity vs. token identity is pointed out by Johnson (1988, p.22).

> "Note that according to the definition of attribute-value structures given above, each constant element appears only once in an attribute-value structure. This means that in fact there is much more 'value sharing' in the actual attribute-value structure than is indicated in the depictions shown above."

This result is again due to extensionality. There are no "copies," which means that the notations are potentially misleading.[31]

The linguistically more familiar way of looking at feature structures is *graph-theoretic*. Thus, the "usual way to conceptualize a feature structure is as a labeled rooted directed graph" (Carpenter 1992, p.36; cf. Shieber 1986, p.20). The AVM in (57) can be translated into the graph in (64).

(64)



Here, each node (reflexively) "dominates" a feature structure. It is important to note that not only nodes but also edges are labeled. In that respect, features structures by far exceed representation of the ID-relation. Take the following definition by Carpenter (1992, p.36).

(65) A *feature structure* over TYPE and FEAT is a tuple $F = \langle Q,q',\theta,\delta \rangle$, where:
- Q: a finite set of *nodes* rooted at q'
- $q' \in Q$: the *root* node
- $\theta: Q \rightarrow$ TYPE: a total node *typing* function
- δ: FEAT$\times Q \rightarrow Q$: a partial *feature value* function

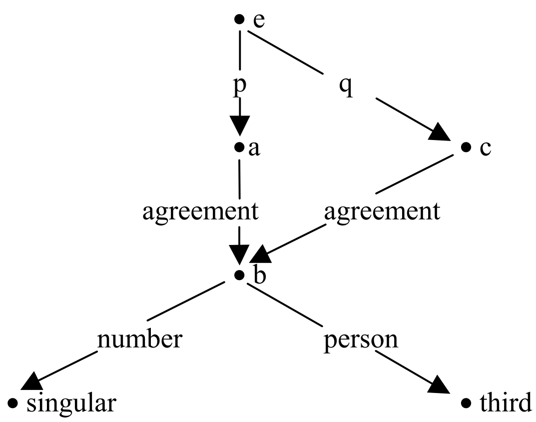
let $\mathbb{F}$ denote the collection of feature structures.

For our purposes we can ignore the difference between types and features underlying the sets TYPE and FEAT.[32] Now, structure sharing arises from relaxing the Single

[31] This question will recur in section 3.3.2 below.

Mother Condition, (8) above, a possibility freely available, given that we are not dealing with trees. Thus, (60) would receive the minimally different representation in (66).[33]

(66)



With this rudimentary understanding of structure sharing, we can turn to some linguistic analyses couched in such terms. The question, of course, arises whether any *theoretical* claims hinge on the use of structure sharing or whether we are merely dealing with a *convenient* linguistic tool (cf. Shieber 1986, chapter 4). Pollard&Sag (1994, p.19) have been very clear on this.

> "It is not going too far to say that in HPSG structure sharing is the central explanatory mechanism, much as move-α is the central explanatory mechanism in GB theory; indeed the relationships between fillers and traces, between 'understood' subjects and their controllers, between pronouns and their antecedents, between 'agreement sources' and 'agreement targets,' and between the category of a word and the category of its phrasal projections will all be analyzed as instances of structure sharing."

For our narrowly minimalist purposes, we can concentrate on the treatment of movement dependencies in attribute-value systems. Consider first a (simplified) LFG-analysis of raising, as discussed in Johnson (1988). In Chomskyan generative syntax, raising is taken to be an instance of A-movement, leaving a trace, as in (67a). LFG, on

[32] As indicated by Pollard&Sag (1987, p.27), there is a close formal connection between feature structures and "finite state automata." Thus, Q corresponds to the set of "states" and δ to the "transition function" (cf. Partee et al. 1993, chapter 17).

[33] Feature structures are often required to be acyclic (Shieber 1986, p.20). For formal discussion of how to add this property, see Johnson (1988, p.50ff) and Carpenter (1992, chapter 5). Note also that "[i]n terms of implementation, it turns out to be more expensive to eliminate cycles than to allow them" (Carpenter 1992, p.35).

the other hand, takes the subject NP to be base-generated in its surface position, as shown in (67b).

(67) a. $[_{IP}$ Mary$_i$ seems $[_{IP}$ t_i to sleep]]
b. $[_{S}$ $[_{NP}$ Mary] $[_{VP}$ seems $[_{VP}$ to sleep]]]

Thus, as far as "c-(onstituent) structure" goes, LFG recognizes only one constituent, *[$_{NP}$ Mary]*, unambiguously immediately dominated by S. The semantically relevant information that Mary is the argument of *to sleep* is encoded in terms of so called "f-(unctional) structure." The f-structure associated with (67b) is given in (68).[34]

(68)

$$\left(\begin{array}{ll} \text{subj} = & {}_{\textcircled{1}}\left[\begin{array}{l} \text{agr} = \left[\begin{array}{l} \text{pers} = \text{3rd} \\ \text{num} = \text{sg} \end{array}\right] \\ \text{pred} = \text{Mary} \end{array}\right] \\ \text{comp} = & \left[\begin{array}{l} \text{subj} = \textcircled{1} \\ \text{pred} = \text{sleep} \end{array}\right] \\ \text{pred} = \text{seem} & \end{array}\right)$$

Value-sharing, indicated by coindexation, guarantees that the feature structure associated with the NP *Mary* assumes the subject function for both S and the minimal VP. Clearly, this analysis presupposes the LFG-approach to grammatical functions, which takes them to be primitives of the theory represented at the level of f-structure. This additional representation lends more flexibility to the use of structure-sharing beyond the modeling of movement chains.

Let us turn to A'-movement next. This time we'll focus on a simplified HPSG analysis.[35] HPSG has fully integrated constituent structures into feature structures, encoding the ID-relation in terms of the attribute DTRS ("daughters"). Combining this

---

[34] Adapted from Johnson (1988, p.17). Pollard&Sag (1994, p.4) likewise assume that "[. . .] there is no need to posit an actual constituent [. . .]" in the case of raising.

[35] For an LFG analysis of unbounded dependencies in terms of f-structures, see Kaplan&Zaenen (1989). There, an extracted element is assigned its grammatical function by means of a "path equation" like (i) (Kaplan&Zaenen 1989, p.27).
(i) ($\uparrow$ TOPIC) = ($\uparrow$ COMP* OBJ)
This roughly says about a fronted ("topicalized") constituent that its GF, OBJ, can be found embedded within an arbitrary number of complements. Unboundedness is thus guaranteed by Kleene closure (*).

kind of attribute with the mechanism of structure sharing, one could translate an MC/MD-graph like the one in (32) above into the AVM below.[36]

(69)
$$\left[\begin{array}{l} \text{F-DTR} = {}_{\textcircled{1}}2 \\ \text{S-DTR} = \left[\begin{array}{l} \text{F-DTR} = 4 \\ \text{S-DTR} = \left[\begin{array}{l} \text{F-DTR} = \textcircled{1} \\ \text{S-DTR} = 7 \end{array}\right]_{5} \end{array}\right]_{3} \end{array}\right]_{1}$$

However, the relation between F-DTR of 1 and F-DTR of 5 is a “nonlocal“ relationship. While this is typical for the output of operations like move-α, it is off-limits for the transformation-less approach pursued by HPSG. Thus, the relation is broken down into a series of local licensing steps. For this purpose, each intermediate projection is provided with an attribute SLASH, linking the positions of the dependency by means of some additional structure sharing. The required additions to (69) are given in rough form in (70).

(70)
$$\left[\begin{array}{l} \text{SLASH} = \varnothing \\ \text{F-DTR} = {}_{\textcircled{1}}2 \\ \text{S-DTR} = \left[\begin{array}{l} \text{SLASH} = \{\textcircled{1}\} \\ \text{F-DTR} = 4 \\ \text{S-DTR} = \left[\begin{array}{l} \text{SLASH} = \{\textcircled{1}\} \\ \text{F-DTR} = \textcircled{1} \\ \text{S-DTR} = 7 \end{array}\right]_{5} \end{array}\right]_{3} \end{array}\right]_{1}$$

Well-formedness conditions can now be imposed locally by means of ID-schemata, much in the spirit of the original context-free phrase structure rules of GPSG (Gazdar et al. 1985), which replace move-α.[37] Of course, given the property of structure sharing,

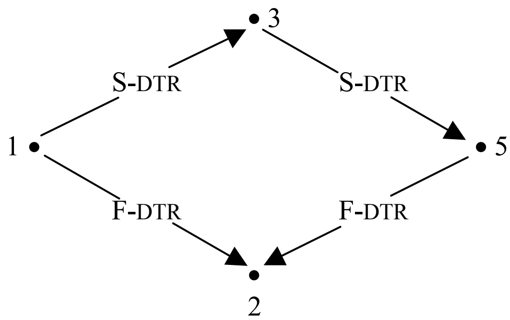
---

[36] For expository purposes, I use binary branching structures, where F-DTR and S-DTR stand for “first“ and “second“ daughter respectively. See Blackburn&Meyer-Viol (1996) for such an approach to graphs. As is well known, GPSG and HPSG advocate the separation of ID and LP relations, i.e. ID schemata do not impose any linear ordering on constituents.

[37] SLASH is set-valued in order to allow multiple extraction. If lists are used instead, constraints like “nestedness“ can be implemented by imposing stacking conditions like “first-in-last-out“ (cf.

which, as we already established, involves token identity, it is less clear what “locality“ really means from a representational perspective. Consider the graph in (71), which translates the essential relations from (70).

(71)



There is no general formal requirement that structure sharing nodes be adjacent in the resulting graph.[38] Rather, the SLASH-mechanism seems to be a special device for dealing with unbounded dependencies in a linguistically satisfactory way. In fact, the amount of structure sharing alotted to A'-dependencies in HPSG is less than what I made it look like so far. Instead of sharing full-fledged constituents, dependencies are formed only wrt “local features,“ i.e. formal features like *category* and *Case*, as well as semantic features. Phonological features, for one thing, are only available at the surface position. Thus, dependencies involve two constituents, namely, a “filler“ and a trace (a.k.a. “gap“), which coincide wrt local features. Structure sharing is pushed to the sub-constituent level.[39] Technically, traces look roughly as follows.

(72)

$$\begin{bmatrix} \text{PHON} = \langle\ \rangle \\ \text{LOCAL} = ① \\ \text{SLASH} = \{①\} \end{bmatrix}$$

These entities introduce a phonologically empty constituent whose local features are in need of identification with those of the filler, which are to be found somewhere along the projection path of the SLASH-feature. In addition, these features can be constrained

---

Pollard&Sag (1994, p.169fn.9). The formal background for the SLASH mechanism is provided by “indexed grammars“ (Partee et al. 1993, p.534ff; Wartena 2000). Note also the close formal connection between SLASH-feature percolation, the Connected Ancestor Condition on MC/MD-graphs and the C-command Condition on Move.

38 See Pollard&Sag‘s (1994, p.17) analysis of the pronoun *she*, where such a putative requirement would have been violated.

39 Compare this approach to the idea that feature-movement is more fundamental than constituent movement advocated in Chomsky (1995a, chapter 4). See also section 2.7.4 above.

by the local syntactic context in terms of subcategorization by a head. By way of an example, (73) and (74) provide a rough analysis of topicalization.

(73) $[_S \text{ Kim}_i \, [_S \text{ Dana } [_{VP} \text{ hates } t_i \,]\,]\,]$

(74)
```
( PHON = <Kim,Dana,hates>
  SLASH = ∅
  LOCAL = ( F-DTR = [ PHON = <Kim>
                      SLASH = ∅
                      LOCAL = [ CAT = NP
                              ① CASE = Acc ] ]
            S-DTR = ( PHON = <Dana,hates>
                      SLASH = {①}
                      LOCAL = ( F-DTR = [ PHON = <Dana>
                                          SLASH = ∅
                                          LOCAL = [ CAT = NP
                                                    CASE = Nom ] ]
                                S-DTR = ( PHON = <hates>
                                          SLASH = {①}
                                          LOCAL = ( F-DTR = [ PHON = <hates>
                                                              SLASH = ∅
                                                              LOCAL = CAT = V ]
                                                    S-DTR = [ PHON = <>
                                                              SLASH = {①}
                                                              LOCAL = ① ] ) ) ) ) )
```

Pollard&Sag (1994, section 9.5) and Sag&Fodor (1994) discuss a traceless alternative to this kind of analysis. Accordingly, no node or terminal at all is used in the position of the gap. Instead, for complement extraction the selecting head is modified by a lexical rule to the effect that the SLASH-feature will be introduced by that head directly. The usual subcategorization constraints can be put on the value of the SLASH-feature inside that lexical item.[40]

Close attention to further technical detail of such rival approaches would, of course, be useful in analyzing minimalist syntax, potentially yielding answers to the questions discussed in section 1.1.[41] However, in compliance with the hypotheses developed in

[40] Subject- and adjunct extraction require additional refinements, discussion of which would carry us too far afield (See Pollard&Sag 1994, section 9.5).

[41] Thus, there may be further "lower" levels of implementation to be taken into account. John Frampton (p.c) has directed my attention to the interpretation of links as "pointers," common in computer science. Implementation of pointers usually requires highly structured objects, located at an address and comprising a "content field" at which some information can be stored as well as "pointer fields" that contain the addresses of further objects of the same kind (cf. Cormen et al. 1990, p.209-213, and Karttunen&Kay 1986). The idea of addressing will show up later in section 3.3.3 as an ingredient of implementing MC/MD in minimalist syntax. Note, incidentally, that

section 2, my foremost task is to show how MC/MD can be added to minimalist syntax and at what cost. Suffice it to say here that there are close links between my proposal and the HPSG approach to (unbounded) dependencies on the basis of structure sharing. To the extent that I'm successful, I will be able to rely on a number of arguments developed within HPSG in defending the use of MC/MD instead of traces or copies.[42]

developing parsing techniques on the basis of copying and “unification“ operations, Karttunen& Kay (1986) come to the conclusion that “the amount of computational effort that goes into producing these copies is much greater than the cost of unification itself“ (Karttunen&Kay 1986, p.6). Thus, in order to save effort they propose to “minimize copying by allowing graphs share common parts of their structure“ (Karttunen&Kay 1986, p.7). Similar views are expressed by Pereira (1986). For another complexity result for copying, see Rogers (1998). If what I have tried to indicate in section 1.1 is along the right track, namely, that memory expenditure of grammar implementations might have an impact on how “perfect“ language is, claims like the ones above might have some importance. Again, I refrain from further speculation here.

42 Cf. Sag&Fodor (1994).

3.3 DoIC or DoID?

In the present section, I develop “MC/MD-syntax,“ a variant of a minimalist system that generates structures essentially equivalent to a subclass of (rooted) MC/MD-graphs. To the extent that this approach is on the right track, it will provide a basis for disposing of the problems outlined in section 2. In particular, my hypotheses 3 and 5 will be met.

The actual order of events will be as follows. I first introduce one version of MC/MD-syntax, based on the structure-building operation “DoIC“ (3.3.1). I then point out a number of problematic aspects of that system, concerning the individuation of terms and checking resources. These problems, I think, would be shared by Merge-based variants of the system as well (3.3.2). While optimists may want to stick with the DoIC approach, I'll develop another more elaborate one, based on the operation “DoID“ supplemented by explicit indexation/individuation procedures (3.3.3). The latter will, I hope, also satisfy the more skeptically inclined.

3.3.1 DoIC[43]

Let me begin by adopting the concept of numeration, $\mathbb{N}$, as a resource for syntactic derivations in the sense of section 2.6.[44] As before, members of $\mathbb{N}$ are fully specified <FPHON,FFORM,FSEM>-triples, that is, “lexical items“ drawn from the set LEX, the lexicon closed under specification operations. These lexical items serve as basic building blocks for syntactic structures generated in MC/MD-syntax. The first constructive step consists in the definition of “terms,“ the objects to be operated on in the system.

(75) Terms
a. Every lexical item is a term
b. If α is a term and β is a term, then $\{\alpha,\{\alpha,\beta\}\}$ is a term and $\{\beta,\{\alpha,\beta\}\}$ is a term
c. Nothing else is a term

Let me next introduce some terminological and notational conventions.

43 There is a certain amount of overlap between the system developed here and ideas presented in Bobaljik (1995a). The latter paper was brought to my attention by Chris Collins (p.c.) only after much of the work here had been completed. I include some discussion of controversial points below.

44 I'll come to questions of individuation later, that is, I assume for the moment that working with sets is unproblematic. For an approach dispensing with numerations, see Collins (1997).

(76) a. Lexical items are called “elementary terms“
b. Non-elementary terms $\{\alpha,\{\alpha,\beta\}\}$ and $\{\beta,\{\alpha,\beta\}\}$ are abbreviated as $<\alpha,\beta>$ and $<\beta,\alpha>$, respectively
c. For any non-elementary term $<\gamma,\delta>$, γ is the “head“ of $<\gamma,\delta>$, δ is the “completion“ of $<\gamma,\delta>$, and $<\gamma,\delta>$ is a “projection“ of γ, i.e. γ “projects“ $<\gamma,\delta>$[45]
d. Terms are alternatively called “constituents“

So far, nothing has happened that differs from Chomsky (1995a) in any interesting way. However, the core syntactic part of MC/MD-syntax diverges from the version of narrow minimalist syntax analyzed in section 2 in two main respects. First, Merge and Move (and Select) are replaced by a single, hybrid operation that establishes immediate constituency. This operation is called “DoIC.“ Secondly, the notion of “derivational stage“ will be enriched so as to offer additional representational options. These are useful in dealing with “weak movement targets“ in a way that avoids countercyclicity (cf. section 2.7). Let's start with the second issue.

While a derivational stage in minimalist syntax comprises just a (multi-)set of “syntactic objects,“ the current system replaces that set by an IC-(=“immediate constituency“) relation and integrates numerations.

(77) A *derivational stage* Σ is a pair $< \mathbb{N}, IC>$, where
a. $\mathbb{N}$ is a set of (elementary) terms, and
b. IC is a binary relation on terms

The definition of “derivation“ will be carried over from section 2.6.2.

(78) A *derivation* is a sequence of derivational stages $\Sigma_0, \ldots, \Sigma_n$ such that for each *i* $(0 < i \leq n)$, Σ_i is the outcome of exactly one syntactic operation applied to Σ_{i-1}.

At Σ_0, IC is empty. Derivational stages are “udated“ by the operation “DoIC.“[46]

(79) *DoIC(α,β)* is a syntactic operation updating a derivational stage Σ (= $<\mathbb{N}, IC>$) to a derivational stage Σ' (= $<\mathbb{N}', IC'>$), such that
a. α and β are terms from $\mathbb{N} \cup$ *dom*(IC) $\cup$ *ran*(IC), $\alpha \neq \beta$,[47] and
b. $\mathbb{N}' = \mathbb{N} - \{\alpha,\beta\}$, and
c. $$IC' = \begin{cases} IC \cup \{ <\alpha,<\alpha,\beta>>, <\beta,<\alpha,\beta>> \}, \text{ if } \alpha \text{ projects} \\ IC \cup \{ <\alpha,<\beta,\alpha>>, <\beta,<\beta,\alpha>> \}, \text{ if } \beta \text{ projects} \end{cases}$$

---

45 The issue of labeling is taken up again in section 3.3.3.
46 Definitions (77) and (79) will receive an addition below.
47 Declaring the non-identity of α and β may be made superfluous by a strict version of feature-checking.

(79b), which captures the function of Select (see section 2.6.2), successively empties the numeration and will be trivially satisfied as soon as $\mathbb{N}$ is empty. (79c) encodes the common denominator of Merge and Move, i.e. that they give rise to a new term and thus to new "immediate constituency" (=IC-) relations.

In terms of IC, we can also define the various dominance relations familiar from section 3.2.

(80) Immediate Dominance
$ID =_{def} IC^{-1}$

(81) Proper Dominance
$PD =_{def} (IC^{+})^{-1}$

(82) Dominance
$D =_{def} (IC^{*})^{-1}$

Also, three additional preparatory conventions should be added here for convenience.

(83) a. An application of DoIC can be abbreviated as DoIC(α,β) = $<\alpha,\beta>$, where $<\alpha,\beta>$ is the new term introduced into *ran*(IC).
b. The input arguments of DoIC are taken to be ordered, such that the first one invariably determines the head of the resulting term, i.e. DoIC(α,β) = $<\alpha,\beta>$ (not $<\beta,\alpha>$)[48]
c. Members from IC are also called "links"

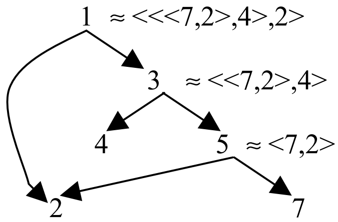
Let me now explain how MC/MD is brought about in this system. Recall from the discussion in section 2.7 that, instead of copying, I want to allow individual constituents to be immediate constituents of two or more other constituents, in relaxation of the Single Mother Condition (cf. 3.2). In fact, nothing prevents a term from being an argument of DoIC more than once during a derivation. Likewise, there is nothing wrong with an IC relation that contains both $<\alpha,\beta>$ and $<\alpha,\gamma>$. That is, there is nothing wrong with an IC relation that allows "multiconstituency," α being an immediate constituent of both β and γ, i.e., α being linked to both β and γ.

The derivation in (85) illustrates how to generate a structure homomorphic to the MC/MD-graph given in (84).[49]

[48] This obviates the need for carrying along the disjunction from (79c) through later definitions.

[49] In order to keep notation of IC relations legible, 5, 3, and 1 replace the actual complex terms in (85), whose internal make-up is indicated in (84). I keep the orientation of edges in the illustrative MC/MD-graphs from dominating node/term to dominated node/term, thus actually picturing not IC but IC^{-1}, i.e. ID.

(84)



(85) a. $\Sigma_0 = < \mathbb{N}_0$ (= { 2, 4, 7 }), IC_0 (= ∅) >
b. *DoIC*(7,2) = <7,2>
c. $\Sigma_1 = < \mathbb{N}_1$ (= { 4 }), IC_1 (= { <2,5>, <7,5> }) >
d. *DoIC*(5,4) = <5,4>
e. $\Sigma_2 = < \mathbb{N}_2$ (= ∅), IC_2 (= { <4,3>, <5,3>, <2,5>, <7,5> }) >
f. *DoIC*(3,2) = <3,2>
g. $\Sigma_3 = < \mathbb{N}_3$ (= ∅), IC_3 (= { <2,1>, <3,1>, <4,3>, <5,3>, <2,5>, <7,5> }) >

The crucial step is (85f), where DoIC applies to term 2 for a second time, making IC_3 of (85g) contain both <2,5> and <2,1>. Thus, a link, or multiconstituency, has been defined without appeal to operations like Move, Copy, or Form Chain.

I'll now turn to the main conditions on DoIC, keeping the system as close to minimalist syntax as possible. Let's start by comparing the structures generated by DoIC to rooted MC/MD-graphs (cf. section 3.2), the definition of which I repeat for convenience.

(86) A *rooted MC/MD-graph* is a pair <N,ID>, where
N is a finite set, the set of nodes,
ID is a binary relation on N, the immediate dominance relation,
and such that
(1) ID^+ is irreflexive (Acyclicity)
(2) $(\exists x \in N)(\forall y \in N)[<x,y> \in ID^*]$ (Single Root Condition)

Clearly, there is a one to one correspondence between nodes and terms, so $N \approx \mathbb{N} \cup$ *dom*(IC) ∪ *ran*(IC). Likewise, ID ≈ IC (cf. (80)). However, given the dynamics of derivations, the Single Root Condition will temporarily have to be violated. This is most clearly the case at Σ_0, where IC is still empty. Thus, as in minimalist syntax (cf. sections 2.5 and 2.6), intermediate stages of a derivation may contain "forests," i.e. graphs with more than one root. The Single Root Condition will only be imposed on the outcome of a derivation. Adapting the corresponding condition from Chomsky (1995a, p.226), I postulate the following constraint on translatability.[50]

(87) SYN = Σ (=<$\mathbb{N}$, IC>) is translatable into PHON/SEM, only if
$(\exists \alpha \in ran(IC))(\forall \beta \in \mathbb{N} \cup dom(IC) \cup ran(IC))[<\beta,\alpha> \in IC^*]$

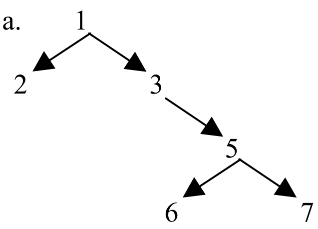
[50] This definition will have to be slightly revised, given an enrichment of Σ introduced below.

It follows from (87) that $\mathbb{N}$ has to be empty at the point where SYN is "spelled-out," i.e. translated into PHON and SEM (cf. section 2.6.1 above). Irreflexivity of IC$^+$, i.e. "Acyclicity," is guaranteed by the peculiar "growth" property of DoIC.[51] Thus, output terms simply collect the input terms into a complex term. It is clear, to mention just the simplest case, that the elementary term α is made an immediate constituent of $\langle\alpha,\beta\rangle$ via DoIC(α,β) by definition, whereas making $\langle\alpha,\beta\rangle$ an immediate constituent of α is impossible, given that, trivially, $\langle\alpha,\beta\rangle$ neither is the head nor the completion of α. No series of additional applications of DoIC can overturn that asymmetry. Thus, not only IC but also IC$^+$ must be irreflexive.[52]

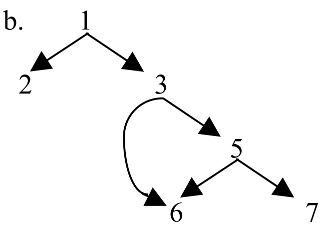
One property distinguishing the structures generated by DoIC from rooted MC/MD-graphs is binary branching. Thus, due to the binary nature of DoIC, no term can immediately dominate more than two terms. Assume DoIC(α,β) = $\langle\alpha,\beta\rangle$. Clearly, no application of DoIC to γ ($\neq \alpha \neq \beta$) can yield $\langle\alpha,\beta\rangle$ again, γ being neither the head nor the completion of $\langle\alpha,\beta\rangle$. As a result, $\langle\gamma,\langle\alpha,\beta\rangle\rangle \notin$ IC.[53] Rooted MC/MD-graphs, on the other hand, include no restriction on branching. We can conclude that the DoIC-system supplemented with condition (87) invariably outputs binary-branching rooted MC/MD-graphs.

Turning to the "dynamic" properties of DoIC, a similar line of reasoning can show that a ban on countercyclicity is an automatic consequence of the system. Thus, consider the hypothetical counter-cyclic transitions in (88) and (89).

(88) a.



b.



[51] I'll keep the following remarks informal, indicating ways of proving the results arrived at in footnotes only.

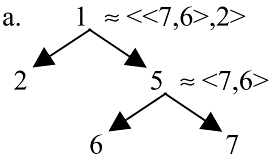
[52] Formally, this result could be derived by assigning each term α a "height," $|\alpha|$, as follows (cf. the notion of "term record" in section 2.6.2).

(i) a. $|\alpha| = 1$ b. $|\langle\alpha,\beta\rangle| = 2$ c. $|\langle\langle\alpha,\beta\rangle,\gamma\rangle| = |\langle\gamma,\langle\alpha,\beta\rangle\rangle| = 3$ d. . . .

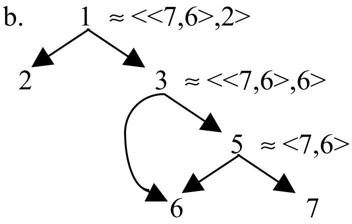
Given the definition of DoIC, for any pair of terms $\langle\alpha,\beta\rangle$, if $\langle\alpha,\beta\rangle \in$ IC then $|\alpha|<|\beta|$. Thus, $\langle\langle\alpha,\beta\rangle,\alpha\rangle \notin$ IC. In general, IC must be irreflexive due to the irreflexivity of $<$, and, given the transitivity of $<$, IC$^+$ must be irreflexive as well.

[53] It is a subtle question whether DoIC could apply twice to the same pair of terms α and β, yielding $\langle\alpha,\beta\rangle$ each time, and whether that could be interpreted as a quaternary branching structure. Technically, this will be excluded by the "Ancestor Condition" on DoIC introduced below. The deeper question concerning identity of terms and the principle of extensionality will be addressed in section 3.3.2. Rigorous conditions on checking (cf. section 3.4.2 and 3.5) could void this issue.

(89) a.

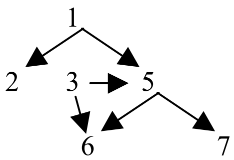


b.



(88) is immediately ruled out, due to the binary nature of DoIC. Thus, unary branching, as required for (88a), i.e. entering <5,3> into IC without there being another term α such that <α,3> ∈ IC as well, is impossible. In (89), on the other hand, the problem is how to relate terms 1 and 3. At stage (89a), <2,1> ∈ IC and <5,1> ∈ IC. The impossibility of ternary branching then precludes the addition of <3,1> by DoIC. Also, there is no way to remove <5,1> from IC. Likewise, at stage (89b) <5,3> ∈ IC and <6,3> ∈ IC, so it is impossible to add <1,3> for the same reason.[54] In that respect, (89b) misrepresents the effect of DoIC(5,6) on (89a). The actually resulting graph would be (90).

(90)



In fact, the kind of ban on countercyclicity built into MC/MD-syntax can be made explicit in terms of the following principle (cf. section 2.5.2).

(91) Strictest Cycle Condition
Every syntactic operation creates a root node

I take the Strictest Cycle Condition to be a sufficient realization of my hypothesis 3.

So far, there is no means to rule out the curious "sideward growing" structure in (90), as long as it occurs at an intermediate derivational stage. What's more, applying DoIC(1,3) to (90) would result in a graph that even satisfies the Single Root Condition. However, intermediate structures like (90) will be eliminated by the condition that replaces the minimalist c-command requirement on Move. Recall from section 3.2 that Barker&Pullum (1990) introduce the Connected Ancestor Condition (CAC), in order to transfer the formal command properties of trees to MC/MD-graphs. I repeat the CAC here.

---

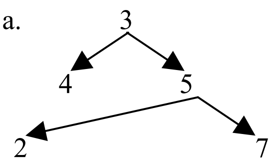
54 See Watanabe (1995) for a similar argument based on banning "redefinition of terms." Again, formally it is the "growth" property of DoIC that prevents terms 1 and 3 from being ordered wrt IC, given that |1| = |3| = 3.

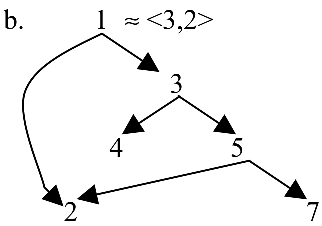
(92) Connected Ancestor Condition (CAC)
$(\forall x,y,z \in N)[\ (\langle x,z\rangle \in ID \wedge \langle y,z\rangle \in ID) \rightarrow (\langle x,y\rangle \in ID^* \vee \langle y,x\rangle \in ID^*)\ ]$

Accordingly, all the ancestors of a given node must be related to each other in terms of dominance. Given that the multiconstituency links created by DoIC are intended to replace movement into c-commanding position, a condition similar to the CAC is required. In order to simulate the derivational character of minimalism as closely as possible, I introduce the following "Ancestor Condition" on Do IC.

(93) Ancestor Condition on DoIC
$\forall\alpha,\beta$, DoIC(α,β) can only apply if
a. $\neg\exists\gamma[\ \langle\alpha,\gamma\rangle \in IC\ ]$, and
b. $\forall\delta[\ \langle\beta,\delta\rangle \in IC \rightarrow \langle\delta,\alpha\rangle \in IC^*\ ]$

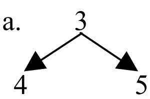
(93a) says that the projecting term must be undominated for DoIC to be applicable. This rules out the transition from (89a) to (89b)/(90), regardless of whether 5 or 6 projects. In welcome transitions like (94), (93a) enforces what in the language of Move is described as "the target projects."[55]

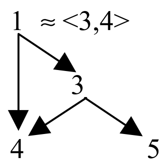
(94) a.


b.


Thus, term 1 in (94b) could not be a projection of 2, i.e. $\langle 2,3\rangle$, given that $\langle 2,5\rangle \in$ IC at stage (94a).

(93b) then effectively requires "connected ancestors." As just stated, in (94), our role-model for a well-behaved transition, $\langle 2,5\rangle \in$ IC at (94a). Therefore, $\langle 5,3\rangle$ must be in IC*, given that 5, the immediate ancestor of 2, must be connected to the future additional immediate ancestor of 2, i.e. 1. This "connection" has to include 3, which is the projecting node α under (93b). In fact, $\langle 5,3\rangle \in$ IC, so (93b) is satisfied. Ordering ancestors in terms of IC*, not IC^+, allows DoIC to also apply in trivial cases such as (95a), where there is just one such ancestor, namely, 3.

(95) a.


b.


[55] Cf. section 2.7.1 for the role of this assumption in guaranteeing uniform chains. See Chomsky (1995a, 4.4.2) for discussion of whether the result is derivable.

Now, (93b) is crucial in ruling out "sideward movement," the final case of movement into non-c-commanding position not yet banned.[56] Consider transition (96).

(96)

Assuming that 2 projects, i.e. $1 \approx \langle 2,4 \rangle$, (93a) is satisfied. However, $\langle 4,3 \rangle \in$ IC and $\langle 3,2 \rangle \notin$ IC*, in violation of (93b). This blocks (96). Note that sideward movement is a way of moving into a non-c-commanding position without violation of the Strictest Cycle Condition. This makes (93b) a necessary addition to MC/MD-syntax.

Let me here provide a very brief and general comparison with of the DoIC-system so far with the basic building blocks of Chomskyan (1995a) minimalist syntax. The minimalist operation Merge is the trivial case under (93), that is, it corresponds to a situation where neither α nor β of DoIC(α,β) is dominated when DoIC applies. The simulation of Move, on the other hand, requires the Ancestor Condition to do a non-trivial job, ruling out what would correspond to movement into non-c-commanding position. It is therefore necessary to emphasize that something like the Ancestor Condition is stipulated in minimalist syntax as well, in order to derive the desired properties for movement and chains. Most explicitly this is stated in the following passage.

> "A chain CH = (α, t(α)) formed by Move meets several conditions, which we take to be part of the definition of the operation itself. One of these is the C-Command Condition: α must c-command its trace, so that there cannot be an operation that lowers α or moves it "sideways;" [. . .] these conditions are part of the definition of the algorithm C_{HL}" (Chomsky 1995a, p.253f).[57]

As amply discussed in section 2, Chomsky (1995a) exceptionally allows two types of counter-cyclic movement, namely, covert movement and X°-movement. Given that the Strictest Cycle Condition, (91), holds in MC/MC-syntax by virtue of the definition of DoIC, these types of movement cannot be captured. I argued in section 2.7 that this is a welcome result for covert movement and subsumed this under hypothesis 3. Consequences for X°-movement will need some reconsideration later (cf. section 3.4.2).

Also, recall from section 2.6 that treating Merge as a "module" of Move in minimalist syntax went along with the problematic distinction between *terms* and *syntactic objects*. That distinction is rendered superfluous by the hybrid operation DoIC,

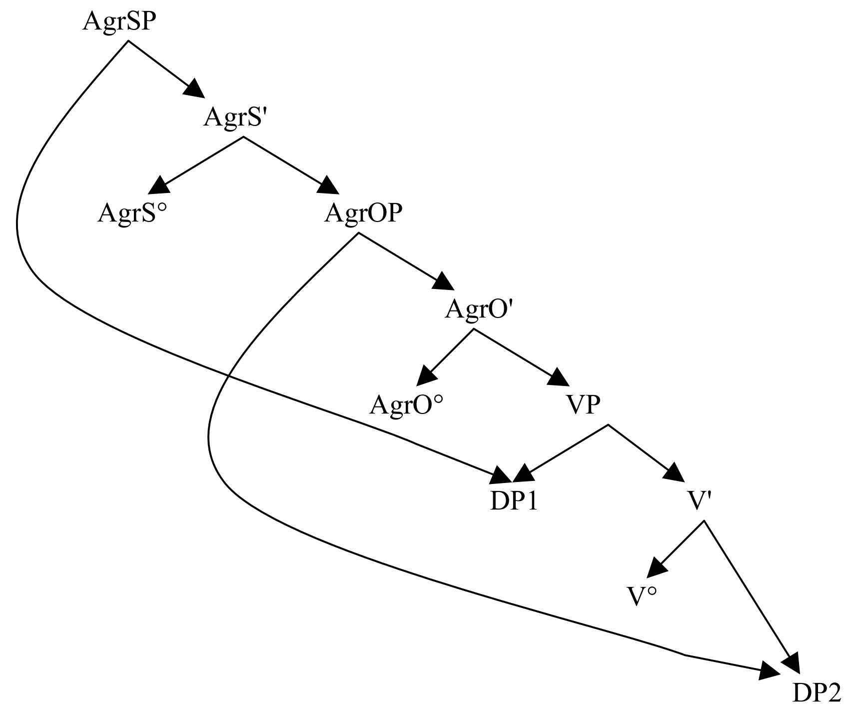
[56] "Lowering" is ruled out as a subcase of countercyclicity (cf. section 3.2). Note that Nunes (1995, 2001) bases an interesting treatment of parasitic gaps on sideward movement. See also Bobaljik (1995) and Bobaljik&Brown (1997), for an application of sideward movement in treating X°-movement.

[57] Collins (1997) explores the possibility of deriving cyclicity from Kayne's LCA at PF. If feasible such a result can be implemented in MC/MD-syntax since the crucial c-command relation is definable (see section 3.4.1).

which invariably applies to *terms* as characterized in (75). It thus looks as if we are on the right track toward meeting the goals set in section 2.

Consider next the question of covert movement again, taking a look at an MC/MD-graph like (97), repeated from section 2.7.3.

(97)



The standard "covert movement" approach (e.g. the treatment of object-placement in English) orders the creation of links for DP1 and DP2 in terms of derivational stages. Thus, DP1 moves before Spell-Out while DP2 does so after Spell-Out. However, the latter requires a counter-cyclic operation and is thus off-limits here. As discussed in 2.7.3, the (obvious) alternative is to create all links in cyclic fashion. It then remains to be decided for multidominated constituents, which link is π-relevant, i.e. which link survives translation into PHON. Clearly, this decision depends among other things on the "strength" of features checked by DP1 and DP2 against AgrS° and AgrO° respectively. We can thus distinguish a "strong link" in the first from a "weak link" in the second case.

In order to implement this distinction in the DoIC-system, I add a second relation on terms to derivational stages, namely, the "weak immediate constituency"(=WIC-) relation. (98), replacing definition (77), takes care of this.

(98) A *derivational stage* Σ is a triple $\langle \mathbb{N}, IC, WIC \rangle$, where
a. $\mathbb{N}$ is a set of (elementary) terms, and
b. IC is a binary relation on terms, and
c. WIC is a binary relation on terms

The necessary enrichment of DoIC for updating WIC is included in (99), which supersedes definition (79).[58]

(99) *DoIC(α,β)* is a syntactic operation updating a derivational stage Σ (= $\langle \mathbb{N}, IC, WIC \rangle$) to a derivational stage Σ' (= $\langle \mathbb{N}', IC', WIC' \rangle$), such that
a. α and β are terms from $\mathbb{N} \cup dom(IC) \cup ran(IC)$, $\alpha \neq \beta$, and
b. $\mathbb{N}' = \mathbb{N} - \{\alpha,\beta\}$, and
c. $IC' = IC \cup \{ \langle \alpha,\langle \alpha,\beta \rangle\rangle, \langle \beta,\langle \alpha,\beta \rangle\rangle \}$, and
d. $WIC' = \begin{cases} WIC \cup \{ \langle \beta,\langle \alpha,\beta \rangle\rangle \}, & \text{if DoIC}(\alpha,\beta) \text{ involves checking of a weak feature} \\ WIC, & \text{otherwise} \end{cases}$

To illustrate the essentials of this procedure, let me have another look at (97). At the stage where AgrOP is built, DoIC applies to the pair <AgrO',DP2>.[59] This involves the checking of a weak feature. Thus, we add <DP2,AgrOP> to both IC and WIC.[60] At the stage where AgrSP is projected, however, checking between DP1 and AgrS° involves a strong feature. Therefore, <DP1,AgrSP> is added only to IC, not WIC. Clearly, this differential treatment of weak and strong links/positions allows us to block spelling-out constituents in the wrong place. Standard minimalist syntax relies on two things for this purpose. (i) Constituents overtly move to the highest strong position they possess features for. The copies they leave behind on this way get deleted somewhere on the PF-branch of C_{HL}. (ii) Covert movement takes place at a stage where feeding the PF-branch has already occurred. Weak positions are thus trivially invisible at π.

In the DoIC-system, in which all positions are created before SYN gets translated into PHON, translation itself has to achieve the same two things. Thus, assume that PHON involves among other things an IC-relation on (FSEM-less) terms, called IC^{φ}. IC^{φ} can be defined in such a way that it corresponds to $IC^{(\sigma)}$ with the exception that (i) among the strong links of multidominated terms the highest one is chosen, and (ii) weak links are eliminated such that only IC-WIC is feeding translation from SYN to PHON.[61] This can be stated as follows.

(100) $\forall \alpha,\beta, \langle \alpha,\beta \rangle \in IC^{\varphi}$ iff
a. $\langle \alpha,\beta \rangle \in$ IC-WIC, and
b. $\neg\exists\gamma[\ \langle \alpha,\gamma \rangle \in \text{IC-WIC} \wedge \langle \beta,\gamma \rangle \in (\text{IC-WIC})^{+}\ ]$

---

58 In (99c) I have eliminated the disjunction in accordance with notational convention (83b).
59 In the following, I will alternatively use standard constituent labels for terms.
60 Technically, the structures manipulated in MC/MD-syntax may therefore be considered "multigraphs" in the sense of Trudeau (1993, p.24).
61 Cf. Peters&Ritchie (1981, p.9), as discussed in section 3.2.

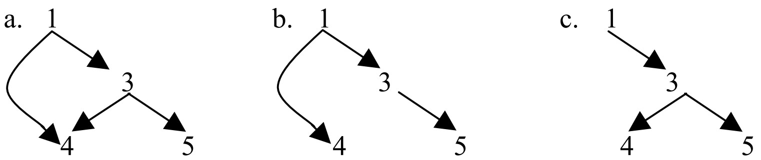
Let's reconsider (97) in the light of this condition and determine the positions of DP1 and DP2 at IC^{φ}. The crucial information is charted in (101). (Relations are restricted to the set Δ = {DP1, DP2}.)

(101) a. IC$\upharpoonright\Delta$ = { <DP2,V'>, <DP1,VP>, <DP2,AgrOP>, <DP1,AgrSP> }
b. WIC$\upharpoonright\Delta$ = { <DP2,AgrOP> }
c. IC-WIC$\upharpoonright\Delta$ = { <DP2,V'>, <DP1,VP>, <DP1,AgrSP> }
d. $IC^{\varphi}\upharpoonright\Delta$ = { <DP2,V'>, <DP1,AgrSP> }

(101d) represents the desired result in that each DP acquires a unique position, DP2 the lower one and DP1 the higher one. <DP2,AgrOP> is filtered out by subtracting WIC from IC. And <DP1,VP>, although being in IC-WIC, is filtered out because there is AgrSP such that <DP1,AgrSP> $\in$ IC-WIC and <VP,AgrSP> $\in$ $(\text{IC-WIC})^{+}$. The latter is due to the fact that { <VP,AgrO'>, <AgrO',AgrOP>, <AgrOP,AgrS'>, <AgrS',AgrSP> } $\subseteq$ IC-WIC also holds.

Let me repeat the general point of this translation procedure into PHON in its crudest way once again. While the hierarchy-based processes of the SYN- (and SEM-) component tolerate multidominance, there is one extra trick required to satisfy the needs of the PHON-component. This trick is easy to formulate in graph-theoretic terms: *Remove redundant edges!* In the DoIC variant of MC/MD-syntax, this corresponds to removing members of IC. Hence, IC^{φ} is the product of subtracting edges from IC, as illustrated in (102).

(102) a. b. c.



The IC relation of (102a) can be translated into IC^{φ} (102b) or IC^{φ} (102c).

Of course, it must be asked whether this procedure could fail in the sense that either no position at all is assigned to a given constituent or more than one. Instrumental in ruling out the first case is the Single Root Condition, as implemented in (87) which, as already observed, ensures that every term enters into some IC-relation, none remaining inside $\mathbb{N}$. In addition, one has to assume that no term is exclusively linked into weak positions. I assume the latter to be a consequence of conditions on feature-checking to be discussed below. Assigning more than one IC^{φ}-link to a given term α is impossible since, due to the Ancestor Condition on DoIC, (93), multiple ancestors of α must be ordered among each other wrt IC*. Thus, condition (100b) will inevitably allow only one of these ancestors to dominate α in IC^{φ}.[62]

[62] Note that this view on SYN>PHON translation is a very rough sketch. For more elaborate alternatives, see e.g. Nunes (1995, 1999) and Uriagereka (1999). Given the unclear internal structure of the PF-branch of C_{HL}, it is an open question how much syntactic information is required there. Thus, eliminating links early, which corresponds to an early deletion of traces, may

Before I compare the MC/MD-approach to covert movement with some rivaling theories, let me give an illustration, how the approach fares wrt such complicated cases as “remnant movement.“[63] In fact, these more complex patterns of constituent displacement cause no trouble for MC/MD-syntax. Consider (103).

(103) [$_{CP}$ [$_{VP}$ t_2 Gelesen]$_1$ hat [$_{IP1}$ [$_{DP}$ das Buch]$_2$ [$_{IP2}$ keiner [$_{I'}$ t_1]]]]
Read has the book no one
“No one has read the book“

Note first, that we are dealing with strong positions only, and that a lot of detail has been omitted. (103) involves the “scrambling“ of *das Buch* out of VP, attaching it to IP1,[64] and subsequent fronting of VP into Spec,CP. Thus, we want <DP,IP1> and <VP,CP> to enter IC$^{\varphi}$, while <DP,VP> and <VP,I'> should be eliminated. For <VP,CP> the desired result is fairly evident from a look at (103) itself. <I',CP> must be a member of (IC-WIC)$^+$. The ordering between VP and IP1, on the other hand, is lost in the trace-based representation. However, the DoIC-system, like the copy theory of movement, is strictly incremental, disallowing such loss of information. Therefore, given that { <VP,I'>, <I',IP2>, <IP2,IP1> } $\subseteq$ IC-WIC holds and thus <VP,IP1> $\in$ (IC-WIC)$^+$, it is <DP,IP1> not <DP,VP> which enters IC$^{\varphi}$, as desired.[65]

Let me now wrap up the discussion of “core“ MC/MD-syntax by looking at three alternative approaches to covert movement that likewise preserve the Strictest Cycle Condition, (87). These are Bobaljik (1995a), Groat&O'Neil (1996), and Stabler (1996). As discussed in section 2.7.2, crucial cases to be dealt with are those which, from a derivational bottom-up perspective, intersperse weak and strong positions to be targeted by one and the same element. The landing site of *WH*-objects in English, as illustrated in (104), provides a prime example.

(104) [$_{CP}$ What$_i$ did [$_{AgrSP}$ John [$_{AgrOP}$ t_i' [$_{VP}$ see t_i]]]]

The DoIC approach links *what* to VP, AgrOP and CP and decides then, on the basis of principle (100), which link survives translation into PHON.

Bobaljik (1995a) assumes basically the same thing. All three positions of *what* are created equal and the decision as to which one is π-relevant is taken later. Explicitly, Bobaljik (1995a, p.53) states that

> “[t]he effects of “covert“ versus “overt“ movement obtain in the phonological component: if the higher copy of a given term is pronounced (lower copies being

be inadequate for treating phenomena like *wanna*-contraction (cf. 2.7.1 and 3.5). Obviously, postponing the reduction procedure would be one option to overcome such problems. I'll briefly return to such matters in sections 3.4.3 and 3.5 below.

63 Cf. Müller (1998) and papers in Alexiadou et al. (eds.) (forthcoming).

64 I sidestep the question how adjunction is properly dealt with here.

65 See Gärtner (1998) for discussion of why remnant movement structures are problematic for the PF-linearization framework defended in Nunes (1995, 1999).

> deleted) then we have the effect of overt movement; if the lower copy is pronounced (higher copies being deleted) then we have the effect of covert movement.“

The common denominator of both approaches, loosely speaking, is a certain “cautious philosophy.“

> “This cautious philosophy can be summarized in the rather cynical slogan: *Always put off till tomorrow what is difficult to decide today*“ (Goldschlager& Lister 1988, p.37).

The two perspectives differ, though, wrt the interpretation of the structures involved. Thus, Bobaljik (1995a) views the output syntactic structure as containing several *copies* of the DP *what*, whereas I interpret the corresponding IC-relations as containing exactly *one* term *what*, being an immediate constituent of each of VP, AgrOP, and CP at the same time.[66]

Going for an early decision on which position should be spelled out is the hallmark of the other two alternatives. Thus, Groat&O'Neil (1996) take movement to leave behind phonological features if it targets a weak position and to carry phonological features along if it targets a strong position.[67] The implementation of this requires a different conception of copying operations. In their own words:

> “[. . .] forming a chain results in copying all syntactic features of the category moved, but does not copy the category's phonological matrix: it either moves it to the new position or fails to move it. This enrichment of Form-chain replaces whatever mechanism is stipulated to mark the tail of a chain as phonologically null“ (Groat&O'Neil 1996, p.125).

---

[66] I'll come back to some subtleties involved in this distinction in section 3.3.2. Both approaches have a tendency toward “trivializing“ the issue of cyclicity and derivational order of events, much like representational theories. See for example Brody (1995), where chains are generated “pre-syntactically.“ It is not entirely clear to me how that system deals with syntactically complex chain links. For further discussion, see Cornell (1999). Another example is Koster's (1987) defense of base-generated S-structures containing empty categories, D-structure and LF being abstractions from S-structure to the extent that they exhibit diverging properties. Dependencies, comprising structures other than filler-gap configurations, are generalized to the abstract relation of “property sharing,“ the conditions on which determine the well-formedness of such representations. Modification of the transformational component, of course, is a running theme in generative grammar. Proposals abound, the frameworks introduced in Abeillé (1993) belonging to the most renowned. Recently, Haider (1997) highlighted a number of “surprising“ properties to be attributed to the output of movement operations, which would be much less surprising if Koster's perspective were correct. Even outside linguistics, movement is not simply taken for granted, as the following quote from elementary physics illustrates: “Questions To Think About: What do we mean by “movement“?“ (Bloomfield 1997, p.3).

[67] Roberts (1998) instantiates a closely related proposal.

However, dealing with cases like (104) poses an extra challenge to their approach. Having created the intermediate weak position without moving FPHON, there is no proceeding from there to the strong position on top, which, among the three positions, requires phonological realization. They therefore seem to be forced to create a second chain, starting form the position where FPHON was left, this time taking phonological features along. (104) would, accordingly, be rendered as (105). (Chains are indicated by lines, transfer of phonological features has been marked as +/-FPHON).

(105) What did John what see what
+FPHON -FPHON

A closely related solution is offered by Stabler (1996). His tree-based formalization of minimalist syntax recognizes three kinds of movement operations. *Move(pi)* displaces an entire subtree containing formal, phonological (p) and interpretable (i) features, leaving behind only an empty subtree. *Move(i)* leaves behind a subtree that possesses only phonological features, the interpretable (i) and formal ones being displaced. *Move*, finally, carries along just formal features, which are then canceled against the ones in the target position. Phonological and interpretable features stay put. No features are ever copied and no chains created.[68]

Again, however, we are back to a decision problem when it comes to (104). Locally, moving *what* into Spec,AgrOP should strand phonological features, applying Move or Move(i). Yet, once formal and phonological features have been thus dissociated, no version of Move could place them both in Spec,CP in the next step.[69] To remedy this situation, Stabler (1996) adopts Chomsky's (1995a) idea of transderivational economy, which decides whether or not to apply Move, Move(i), or Move(pi). The more features one carries along the costlier. Crucially, this economy principle is violable in case it turns out that a locally parsimonious choice fails to satisfy the global context. Thus, Move(pi) can be applied even for the first step to a weak position in (104).[70]

Let me repeat that all three alternatives share with the DoIC-system the ban on counter-cyclic operations. They are thus all compatible with hypothesis 3. However, appeal to copies makes the approaches of Bobaljik (1995a) and Groat&O'Neil (1996) vulnerable to (variants of) the “resource problem“ and the “explosion problem“ discussed in section 2.7. These problems do not arise in frameworks where “nothing is ever copied,“ such as Stabler (1996) and the MC/MD-approach defended here. The latter two systems diverge on whether or not there should be appeal to economy principles in syntax. In line with hypothesis 6 (cf. section 2.8), MC/MD-syntax is designed to make such an appeal superfluous.

Summing up what has been achieved so far, I suggest that a derivational system can be built that closely simulates minimalist syntax as developed in Chomsky (1995a)

68 This procedure seems to require semantic interpretation in parallel with syntactic derivations.

69 In fact, there is no operation that moves just phonological features, so once they are stranded they are frozen in place.

70 See sections 2.7.2 and 2.8 for discussion of the decision problem and economy principles.

while at the same time meeting hypotheses 3 - 6 (cf. section 2).[71] The crucial property of "multiconstituency"/"multidominance" (MC/MD) involved in this undertaking is induced by two main factors.

(i) Derivational stages are enriched such that over and above a numeration they contain both an "immediate constituency"(IC-) and a "weak immediate constituency" (WIC-)relation. These relations replace a (multi-)set of "syntactic objects" in minimalist derivations. The relevant definition, (98), is repeated here as (106).

(106) A *derivational stage* Σ is a triple $\langle \mathbb{N}, \mathrm{IC}, \mathrm{WIC}\rangle$, where
a. $\mathbb{N}$ is a set of (elementary) terms, and
b. IC is a binary relation on terms, and
c. WIC is a binary relation on terms

(ii) A hybrid operation called "DoIC" compounds the workings of Select, Merge, and Move to build up an IC-relation among constituents. It crucially allows individual constituents to become the immediate constituent of more than one larger constituent. This feature is introduced by allowing constituents to be an argument of DoIC, and thus acquire an immediate ancestor, more than once. (99), the definition of DoIC, is repeated here as (107).

(107) *DoIC*(α,β) is a syntactic operation updating a derivational stage Σ (= $\langle \mathbb{N}, \mathrm{IC}, \mathrm{WIC}\rangle$) to a derivational stage Σ' (= $\langle \mathbb{N}', \mathrm{IC}', \mathrm{WIC}'\rangle$), such that
a. α and β are terms from $\mathbb{N} \cup \mathit{dom}(\mathrm{IC}) \cup \mathit{ran}(\mathrm{IC})$, $\alpha \neq \beta$, and
b. $\mathbb{N}' = \mathbb{N} - \{\alpha,\beta\}$, and
c. $\mathrm{IC}' = \mathrm{IC} \cup \{ \langle\alpha,\langle\alpha,\beta\rangle\rangle, \langle\beta,\langle\alpha,\beta\rangle\rangle \}$, and
d. $\mathrm{WIC}' = \begin{cases} \mathrm{WIC} \cup \{ \langle\beta,\langle\alpha,\beta\rangle\rangle \}, & \text{if DoIC}(\alpha,\beta) \text{ involves checking of a weak feature} \\ \mathrm{WIC}, & \text{otherwise} \end{cases}$

The compounding of Merge and Move into DoIC makes it possible to reduce the "ontological" distinction among constituents of "syntactic objects" and "terms" in minimalist syntax (cf. section 2.6) to just "terms."

The output of the MC/MD-system are rooted MC/MD-graphs as defined in section 3.2. As in minimalist syntax, intermediate derivational stages allow a multiplicity of graphs to coexist in "forests," the Single Root Condition, (87), repeated in slightly revised form in (108), being a condition only on output.

(108) SYN = Σ (=$\langle \mathbb{N}, \mathrm{IC},\mathrm{WIC}\rangle$) is translatable into PHON/SEM, only if
$(\exists\alpha \in \mathit{ran}(\mathrm{IC}))(\forall\beta \in \mathbb{N} \cup \mathit{dom}(\mathrm{IC}) \cup \mathit{ran}(\mathrm{IC}))[\, \langle\beta,\alpha\rangle \in \mathrm{IC}^* \,]$

[71] I'll come back to H2 in section 3.4.2 and 3.5. I consider H4 and H6 to be met implicitly, given the absence of appeals to "invisibility" and economy principles from MC/MD-syntax.

It follows from the definition of DoIC that the structures manipulated in MC/MD-syntax are binary branching and that derivational transitions observe the Strictest Cycle Condition, (91), repeated in (109).

(109) Strictest Cycle Condition
Every syntactic operation creates a root node

Multiconstituency is restricted to a simulation of movement into c-commanding position. This is guaranteed by the Ancestor Condition on DoIC, (93), repeated as (110), which replaces the minimalist C-Command Condition on Move.

(110) Ancestor Condition on DoIC
$\forall\alpha,\beta$, DoIC(α,β) can only apply if
a. $\neg\exists\gamma[<\alpha,\gamma> \in \text{IC}]$, and
b. $\forall\delta[<\beta,\delta> \in \text{IC} \rightarrow <\delta,\alpha> \in \text{IC}^*]$

(110) forces the immediate ancestors of any given constituent to be ordered in terms of dominance among each other, a property the immediate ancestors of chain links in a chain observing c-command share.

Finally, multidominated constituents are assigned their π-relevant position by a SYN>PHON-translation rule that eliminates all but the highest link arisen from a strong checking relation. The required definition, (100), is repeated in (111).

(111) $\forall\alpha,\beta, <\alpha,\beta> \in \text{IC}^{\varphi}$ iff
a. $<\alpha,\beta> \in$ IC-WIC, and
b. $\neg\exists\gamma[<\alpha,\gamma> \in \text{IC-WIC} \wedge <\beta,\gamma> \in (\text{IC-WIC})^+]$

This completes my introduction of basic MC/MD-syntax. What is to follow will fundamentally preserve the structures and operations defined here. Sections 3.3.2 and 3.3.3 will provide a conceptual shift, which makes addition of a sufficiently powerful indexation device seem advisable.

3.3.2 Extensionality and Graphs as “Pictures“ of Sets

This section discusses the role of set-theoretic assumptions in generating MC/MD-structures. It turns out that a commitment to extensionality is involved. To the extent that this is problematic, a revision of the DoIC-system is called for.

Consider once again the kind of dependency we have been looking at, cast in minimalist set notation. Labeling is still set aside.

(112) $\{ \alpha, \{ \delta, \{ \gamma, \{ \alpha, \beta \} \} \} \}$

In section 2.7, I concluded that, given the copy theory of movement and standard notions of locality, the two occurrences of α in (112) are not meant to be identified in minimalist syntax.[72] This then gives rise to the “resource problem“ and the “explosion problem“ discussed in sections 2.7.2 and 2.7.4. According to my hypothesis 5, an identification of the two occurrences, i.e. appeal to MC/MD, is a proper way out.

Now, set-theoretically there is nothing wrong with some α being a member of more than one set at once, whatever the relation between those sets may be. In fact, if we assume that all objects in (112) are sets, identity of the two occurrences of α follows from the Axiom of Extensionality (Suppes 1972, p.21; cf. Essler 1982, p.45f).

(113) Axiom of Extensionality
$(\forall x)[\, x \in A \leftrightarrow x \in B \,] \rightarrow A = B$

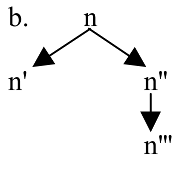
Thus, it seems as if a commitment to set theory in general and the Axiom of Extensionality in particular would allow a direct reanalysis of minimalist structures in terms of MC/MD. (112) can simply be taken to “be“ a rooted MC/MD-graph in which α is both an immediate constituent of $\{\alpha,\beta\}$ and $\{\alpha,\{\delta,\{\gamma,\{\alpha,\beta\}\}\}\}$. The two occurrences of α are merely forced upon us by bracket notation as another instance of “excess of notation over subject-matter“ (Quine 1995/1941[1], p.5).

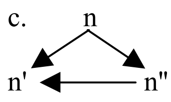
This perspective is basically confirmed if we take a look at an analysis of the relation between sets and graphs provided by Aczel (1988).[73] Accordingly, graphs can be considered “pictures“ of sets, given the following definitions (Aczel 1988, p.4).

(114) a. A *pointed graph* is a graph together with a distinguished node called its point
b. A pointed graph is *accessible* if for every node *n* there is a path $n_0 \rightarrow n_1 \rightarrow \ldots \rightarrow n$ from the point n_0 to the node *n*
c. A *decoration* of a graph is an assignment of a set to each node of the graph in such a way that the elements of the set assigned to a node are the sets assigned to the children of that node
d. A *picture* of a set is an accessible pointed graph (apg) which has a decoration in which the set is assigned to the point

For example, the set-theoretic version of natural number two, given in (115a), can be associated with the “pictures“ (115b) and (115c), on the basis of decorations (116a) and (116b) respectively.

(115) a. $\{ \varnothing, \{ \varnothing \} \}$

b.


c.


[72] See section 3.4.1, for further evidence why this construal of standard minimalist syntax is accurate.
[73] Thanks to Marcus Kracht and Reinhard Blutner for directing my attention to this work, and thanks to Stefan Geschke for discussing some of the issues with me.

(116) a. *dec*(115b) = $\{\langle n''',\varnothing\rangle, \langle n'',\{\varnothing\}\rangle, \langle n',\varnothing\rangle, \langle n,\{\varnothing,\{\varnothing\}\}\rangle\}$
b. *dec*(115c) = $\{\langle n'',\{\varnothing\}\rangle, \langle n',\varnothing\rangle, \langle n,\{\varnothing,\{\varnothing\}\}\rangle\}$

Thus, one and the same set can be given a picture in terms of graphs that do or do not satisfy the Single Mother Condition, i.e. a constituent structure tree, (115b), and a rooted MC/MD-graph, (115c), respectively.

Consider next Aczel's “Corollary 1.1.“ based on “Mostowski's Collapsing Lemma.“ The additional notion of “well-foundedness“ is defined in (117a) (Aczel 1988, p.4f).

(117) a. A graph is *well-founded* if it has no infinite path
b. Mostowski's Collapsing Lemma
Every well-founded graph has a unique decoration
c. Corollary 1.1.
Every well-founded apg is a picture of a unique set

Crucially, uniqueness of the picture relation holds in the direction from graphs to sets, not vice versa. What distinguishes graphs from sets, of course, is the notion of node. Nodes are the constitutive individuals of graphs. Thus, (115b) differs from (115c) among other things by virtue of containing an additional node. As for sets, things depend on how much set theory one is committed to. This is the result we already arrived at above. If the Axiom of Extensionality is adopted, α in (112) and $\varnothing$ in (115a) count as one individual each. In terms of number of individuals, then, (115c) corresponds to (115a) closer than (115b) does.

However, the difficult question is what adopting the Axiom of Extensionality means for (minimalist) syntactic theory. Clearly, this involves a fundamental decision wrt the level of abstraction at which grammatical representations relate to symbolic computation. As already indicated in section 1.1, minimalism invites exploration of such fundamental questions. Of course, it would be pretentious of myself to even think of trying to settle the issue. What I hope my discussion minimally shows is that parts of minimalist syntax are permeated by such difficult questions.[74]

In fact, the stance in Chomsky (1995a) vis-a-vis set theory is ambiguous. As noted in section 2.6.3, the definition of the label of a term seems to presuppose extensionality. Thus, in a labeled term $\{\alpha,\{\alpha,\beta\}\}$ the leftmost occurrence of α, i.e. the label, is taken to be identical to the second occurrence of α, i.e. the (elementary) subterm α. On the other hand, there seems to be an explicit attempt to ban such powerful devices, as the following quote, repeated from section 2.6.1, indicates.

> “Note that considerations of this nature can be invoked only within a fairly disciplined minimalist approach. Thus, with sufficiently rich formal devices (say, set theory), counterparts to any object (nodes, bars, indices, etc.) can readily be constructed from features. There is no essential difference, then, between admitting new kinds of objects and allowing richer use of formal devices; we

[74] For a discussion of extensionality wrt feature-structures, see Carpenter (1992).

assume that these (basically equivalent) options are permitted only when forced by empirical properties of language (Chomsky 1995a, p.381fn.7)."[75]

The power of set-theoretic extensionality can be seen most clearly when the dynamics of feature-checking is taken into account. Thus, consider a derivational transition from (118a) to (118b), establishing a checking relation between α and δ.

(118) a. $\{ \delta, \{ \gamma, \{ \beta, \alpha \} \} \}$
b. $\{ \alpha, \{ \delta, \{ \gamma, \{ \beta, \alpha \} \} \} \}$

As discussed in section 2.7, if checking is purely local, along with δ only one occurrence of α gets substituted. This is shown in (119).

(119) $\{ \alpha', \{ \delta', \{ \gamma, \{ \beta, \alpha \} \} \} \}$

However, if the Axiom of Extensionality holds, (119) is not the correct result of substituting α in (118b). Instead, (120) would be derived.

(120) $\{ \alpha', \{ \delta', \{ \gamma, \{ \beta, \alpha' \} \} \} \}$

Thus, if we decide to consider notation epiphenomenal, we arrive at the MC/MD-solution to the problems outlined in section 2 via set-theoretic extensionality.[76]

It is therefore time to review the DoIC-system in the light of these results. In fact, when it comes to notation, that system is much more inflationary than minimalist set notation. Consider the IC-relation corresponding to (118a), given in (121).[77]

(121) $IC = \{ <\alpha,<\beta,\alpha>>, <\beta,<\beta,\alpha>>, <\gamma,<<\beta,\alpha>,\gamma>>, <<\beta,\alpha>,<<\beta,\alpha>,\gamma>>,$
$<\delta,<\delta,<<\beta,\alpha>,\gamma>>>, <<<\beta,\alpha>,\gamma>,<\delta,<<\beta,\alpha>,\gamma>>> \}$

Crucially, transition to the equivalent of (118b) involves adding $<\alpha,<\delta,<<\beta,\alpha>,\gamma>>>$ to IC in (121), i.e. α must be linked to the root. Thus, multiconstituency for α in the MC/MD-version of (118b) boils down to the coexistence of the two ordered pairs, i.e. "edges," given in (122).

---

75 This, unsurprisingly, is in clear contrast with the view in Chomsky (1975/1955[1], p.107fn.4), where each linguistic level is taken to contain a full set theory to "simplify the constructive task."

76 As indicated in section 2.7.4, the line pursued in Chomsky (2000, 2001) and Frampton&Gutmann (1999) is to get from (118a) to (120) not via (118b) but via (i), i.e. by means of long-distance checking.
(i) $\{ \delta', \{ \gamma, \{ \beta, \alpha' \} \} \}$
This would seem to obviate an appeal to the Axiom of Extensionality. However, see 2.7.4 above for discussion of why this might still not suffice to solve the problems involved in copying.

77 I have arbitrarily chosen the projecting nodes where such a choice is open. Note that the IC-relation of MC/MD-syntax shares this inflationary use of symbols with the phrase markers of Chomsky (1975/1955[1]) and the reduced phrase markers of Lasnik & Kupin (1977). Cf. section 2.3 above.

(122) $\mathrm{IC}'\restriction\{\alpha\} = \{ \langle\alpha,\langle\beta,\alpha\rangle\rangle, \langle\alpha,\langle\delta,\langle\langle\beta,\alpha\rangle,\gamma\rangle\rangle\rangle \}$

Next, if feature-checking continues to mean substitution, all the occurrences of α and δ have to be replaced at one and the same time, as illustrated in (123).

(123) $\mathrm{IC}''\restriction\{\alpha'\} = \{ \langle\alpha',\langle\beta,\alpha'\rangle\rangle, \langle\alpha',\langle\delta',\langle\langle\beta,\alpha'\rangle,\gamma\rangle\rangle\rangle \}$

Thus, for the DoIC-system to successfully implement rooted MC/MD-graphs, an appeal to the Axiom of Extensionality is equally necessary. Only then can the inflation of substitutions be considered a notational epiphenomenon. Again, we have a gap between set-theoretic devices and the intended graph-theoretic "picture." In fact, I already relied on the power of set theory in section 3.3.1, when equating the set of nodes N of an MC/MD-graph with the set $\mathbb{N} \cup$ *dom*(IC) $\cup$ *ran*(IC). Indeed, trivially, *dom*($\mathrm{IC}'\restriction\{\alpha\}$) = {α}, i.e. the IC-relation in (122) gives rise to only one "node" α, which is the desired result.

Note, finally, that the system presented in Bobaljik (1995a) faces exactly the same kind of challenge. There, (118a) would be rendered as an "unordered list," of "terms," as illustrated in (124).

(124) $\alpha; \beta; \{ \alpha, \beta \}; \gamma; \{ \gamma, \{ \alpha, \beta \} \}; \delta; \{ \delta, \{ \gamma, \{ \alpha, \beta \} \} \}$

Transitions are brought about by a single operation, Merge, which applies to two terms on the list and adds a new one. Crucially, the input terms are not replaced. Thus, transition from (118a) to (118b) requires an application of Merge(α,{δ,{γ,{α,β}}}) at stage (124), yielding (125).

(125) $\alpha; \beta; \{ \alpha, \beta \}; \gamma; \{ \gamma, \{ \alpha, \beta \} \}; \delta; \{ \delta, \{ \gamma, \{ \alpha, \beta \} \} \};$
$\{ \alpha, \{ \delta, \{ \gamma, \{ \alpha, \beta \} \} \} \}$

Again, the proper substitutions required for feature-checking would have to rely on the Axiom of Extensionality. However, the explicit reference to multiple inhabitants of the resulting structures as "copies" makes it unlikely that such an appeal is intended.[78]

As already indicated in the opening of section 3.3, I consider the foregoing discussion reason enough to be skeptical of the DoIC-system as it stands. I will therefore proceed to a more elaborate but closely related system, based on the operation

[78] Reformulated in the vocabulary of Chomsky (1995a), Bobaljik (1995a) makes a transition from (118a) to (118b) in terms of the operation "Merge" possible by keeping a record of "syntactic objects" in his structures. Thus, (124) contains a "syntactic object" α which Merge can apply to a second time. As repeatedly noted before, the same transition has to resort to the operation Move in Chomsky (1995a). Move applies to the "syntactic object" { δ, { γ, { α, β } } }, copies one of its terms, namely, α, and "type-lifts" that term to the rank of "syntactic object." Then Merge, as a "module" of Move can apply to { δ, { γ, { α, β } } } and the properly typed copy of α, yielding (118b). This kind of "type-lifting" is implicitly assumed by Nunes (1995) and Collins (1997), who allow the copying of *terms* followed by inserting them into the derivation as *syntactic objects*.

DoID, supplemented with explicit individuation/indexation procedures. I take it that standard minimalist syntax would have to be supplemented with a proper theory of chain structures to be on the safe side wrt extensionality as well.[79] Although the DoID-system does not fully eliminate the role of extensionality, it trivializes that role.

3.3.3 DoID

Let me start by addressing another potentially controversial property of the DoIC-system. Clearly, the IC-relation encodes the edges of the corresponding graph. However, the members of *ran*(IC) do the same thing all by themselves, given their internal structure. Thus, in order to express the fact that α is an immediate constitutent of $<\alpha,\beta>$, the pair $<\alpha,<\alpha,\beta>>$ has to be added to IC. However, the same information is likewise encoded through $<\alpha,\beta>$ alone, given its internal make-up. The DoIC-system could therefore be recast in minimalist notation, if DoIC were replaced by "Do$\in$."[80] An application of Do$\in(\alpha,\beta)$ could then simply generate $\{\alpha,\beta\}$ instead of the prima facie redundant $\in = \{ <\alpha,\{\alpha,\beta\}> , <\beta,\{\alpha,\beta\}> \}$.[81] Of course, this redundancy is due to the nature of terms, which correspond both to the nodes of a graph and the subgraphs these nodes dominate.

The DoID-system introduced in the present section will eliminate this redundancy, replacing terms with so-called "pointers." These will be fully individuated and correspond only to nodes of a graph. Encoding the edges of the graphs will then be the function of a non-redundant ID-relation.

A second shift responds to the issue of substitutions and extensionality raised in the previous section. Thus, lexical items, which introduce the checkable features into derivational stages, will permanently be kept at just one place, the so-called "terminal numeration" ($\mathbb{N}_T$). This domain will be kept apart from an "address domain," i.e. the ID-relation occupied by the "pointers." Conceptually, of course, this is close to the structure-sharing approach, as discussed in section 3.2.1.[82]

As in the DoIC-system, lexical resources are the specified lexical items from LEX. In order to keep them separate from the actual "nodes" of the MC/MD-graph, the DoID-system first generates coindexed counterparts of the lexical items. These counterparts, eventually referred to as "pointers," will not contain any checking resources, or O-features (cf. section 3.1). The MC/MD-graph will then be projected from such pointers by the operation DoID in a way closely analogous to the earlier system.

The indexing procedure, which is designed to respect the principle of inclusiveness (cf. section 2.6.1), is stated in terms of an operation called "Select." This operation inserts lexical items into so-called "(pre-)terminalization structures."

---

79 Cf. section 2.7. For the formal background of such a theory, see Kracht (2001).

80 Thanks to Reinhard Muskens (p.c.) for raising this point.

81 This would, of course, leave the need for dealing with covert movement, i.e. for a technique that replaces WIC.

82 For additional formal background on pointers, see Cormen et al. (1990, p.209-213).

(126) A *(pre-)terminalization structure* (PTS) is an ordered triple $\langle \mathbb{N}_T, \mathbb{N}_P, \mathbb{N}_I \rangle$, such that
- a. $\mathbb{N}_T$ is a set of ordered pairs $\langle x,I \rangle$, *x* a lexical item, *I* a multiset called “index,“ and
- b. $\mathbb{N}_P$ is a set of ordered pairs $\langle B,I \rangle$, *B* a set of categorial features called “bar-level,“ *I* a multiset called “index,“ and
- c. $\mathbb{N}_I$ is a multiset of categorial features called “index“

I will adhere to the following terminology.

(127)
- a. $\mathbb{N}_T$ will be called “terminal numeration“
- b. $\mathbb{N}_P$ will be called “preterminal numeration“

Indexation, brought about by Select(α), uses the categorial feature *cat*(α) of the lexical item α inserted.[83]

(128) *Select(α)* is an operation that applies to α, a member from LEX, and updates PTS to PTS', such that
- a. $\mathbb{N}_T' = \mathbb{N}_T \cup \{ \langle \alpha, \mathbb{N}_I \rangle \}$, and
- b. $\mathbb{N}_P' = \mathbb{N}_P \cup \{ \langle \{cat(\alpha)\}, \mathbb{N}_I \rangle \}$, and
- c. $\mathbb{N}_I' = \mathbb{N}_I \cup [\, cat(\alpha) \,]$

(128a) and (128c) are one explicit version of the individuation of (occurrences of) lexical items, which is required for minimalist syntax as well (cf. 2.6.1).[84] (128b) then extends that method for creating a set of counterparts of lexical items. These are based on just categorial features and thus free of O-features (cf. section 3.1).

Let me go through a sequence of such indexation steps. Assume that *cat*(α) = *a* and *cat*(β) = *b*.

(129)
- a. $\mathbb{N}_T = \varnothing;\ \mathbb{N}_P = \varnothing;\ \mathbb{N}_I = [\ ]$ — b. Select(α)
- c. $\mathbb{N}_T' = \{ \langle \alpha, [\ ] \rangle \}$
  $\mathbb{N}_P' = \{ \langle \{a\}, [\ ] \rangle \}$
  $\mathbb{N}_I' = [\, a \,]$ — d. Select(β)
- e. $\mathbb{N}_T'' = \{ \langle \alpha, [\ ] \rangle, \langle \beta, [\, a \,] \rangle \}$
  $\mathbb{N}_P'' = \{ \langle \{a\}, [\ ] \rangle, \langle \{b\}, [\, a \,] \rangle \}$
  $\mathbb{N}_I'' = [\, a,b \,]$ — f. Select(α)
- g. $\mathbb{N}_T''' = \{ \langle \alpha, [\ ] \rangle, \langle \beta, [\, a \,] \rangle, \langle \alpha, [\, a,b \,] \rangle \}$
  $\mathbb{N}_P''' = \{ \langle \{a\}, [\ ] \rangle, \langle \{b\}, [\, a \,] \rangle, \langle \{a\}, [\, a,b \,] \rangle \}$
  $\mathbb{N}_I''' = [\, a,b,a \,]$

---

83 Multisets are enclosed in brackets, [ ], ∪ will be equally used for set- and multiset-union.

84 See Nunes (1995) for a similar procedure. The movement operation employed there turns out to also incorporate an indexation operation.

These three steps suffice for illustrating the essentials of the procedure. First, due to the multiset-nature of $\mathbb{N}_I$, each application of Select will create a fresh index. The non-trivial step is shown in transition (129e) to (129g), where *a* is inserted into $\mathbb{N}_I$ for a second time. As a consequence, secondly, each inserted lexical item will obtain a unique index, irrespective of whether lexical items derive from identical lexical types or not. In (129), selecting α twice gives rise to two distinguishable occurrences of α, namely, <α,[ ]> and <α,[a,b]>.

The following terminological conventions have to be added at this point.

(130) a. A pair of a lexical item and an index is called “terminal“
b. A pair of a singleton bar-level and an index is called “preterminal“

The indexation procedure so far is responsible for “horizontal“ individuation of terminals and preterminals. The role of the bar-level will be to bring about “vertical“ individuation as well.[85] This is required for generating “pointers,“ i.e. the objects on which MC/MD-graphs are defined in the DoID-system. They replace the “terms“ of the DoIC-system.[86]

(131) Pointers
a. Every preterminal is a pointer
b. If <A,I> is a pointer, then <A¤A,I> is a pointer
c. Nothing else is a pointer

Take for example the preterminal <{a},I>. <{a},I> is a pointer. But then so are <{a}¤{a},I> = <{a,{a}},I> and <{a,{a}}¤{a,{a}},I> = <{a,{a},{a,{a}}},I> and so on. Given the obvious relation to the construction of natural numbers, I allow myself to abbreviate bar-levels by their corresponding digits, i.e. <1,I> for <{a},I>, <2,I> for <{a,{a}},I>, <3,I> for <{a,{a},{a,{a}}},I>, and so on.

Next I define the notions “projection,“ relating pointers among each other, and “terminal head,“ relating pointers to terminals.

(132) For every x,y, such that x (= <A,I>) is a pointer and y (= <B,J>) is a pointer, x is a *projection* of y, iff I = J and B ∈ A.

(133) For every x,y, such that x (= <α,I>) is a terminal and y (= <A,J>) is a pointer, x is the *terminal head* of y, iff I = J.

85 The property of individuation is also fundamental to Higginbotham's (1985b) definition of phrase markers, which “consist of a finite set of (occurrences of) linguistic elements, among which are included the formatives and the categorial symbols [. . .]“ (Higginbotham 1985b, p.88). Appeal to occurrences is meant to render superfluous reference to nodes and a labeling function (cf. Ojeda 1988).

86 ¤ is the set-insertion operator from Manna&Waldinger (1985). Cf. section 2.6.2 above. See also section 2.6.4, for this construction of bar-level marking.

The expression "terminal head of α" will be abbreviated as "$th(\alpha)$." Note that indexation does what labeling does in standard minimalism, i.e. it signals the location of those checking resources which determine the combinatorial properties of the current projection. Again, pointers allow an O-feature-free version of this.[87]

Let me next turn to more familiar ground again, namely, the definition of derivational stages. While the DoIC-system built these in terms of just numerations, here both *terminal* and *preterminal* numerations are employed. The latter two can be "abstracted" from (pre-)terminalization structures as follows.

(134) For every PTS = $\langle \mathbb{N}_T, \mathbb{N}_P, \mathbb{N}_I \rangle$ there is a *reduced (pre-)terminalization structure, RPTS* = $\langle A,B \rangle$, such that A = $\mathbb{N}_T$ and B = $\mathbb{N}_P$.

(135) A *derivational stage* Σ is a triple $\langle \mathbb{R}, ID, WID \rangle$, such that
a. $\mathbb{R}$ is an RPTS, $\langle \mathbb{N}_T, \mathbb{N}_P \rangle$, and
b. ID is a binary relation on pointers, and
c. WID is a binary relation on pointers

As before, then, we need an operation updating derivational stages.[88]

(136) *DoID(α,β)* is a syntactic operation updating a derivational stage Σ (= $\langle\langle \mathbb{N}_T, \mathbb{N}_P \rangle, ID, WID \rangle$) to a derivational stage Σ' (= $\langle\langle \mathbb{N}_T', \mathbb{N}_P' \rangle, ID', WID' \rangle$), such that
a. α (= $\langle A,I \rangle$) and β (= $\langle B,J \rangle$) are pointers from $\mathbb{N}_P \cup dom(ID) \cup ran(ID)$, $\alpha \neq \beta$, and
b. $\mathbb{N}_T' = \begin{cases} \mathbb{N}_T[th(\alpha)'/th(\alpha), th(\beta)'/th(\beta)], & \text{if DoID}(\alpha,\beta) \text{ involves feature checking} \\ \mathbb{N}_T, & \text{otherwise} \end{cases}$
c. $\mathbb{N}_P' = \mathbb{N}_P - \{\alpha,\beta\}$, and
d. $ID' = ID \cup \{ \langle\langle A\text{¤}A,I \rangle,\alpha \rangle, \langle\langle A\text{¤}A,I \rangle,\beta \rangle \}$, and
e. $WID' = \begin{cases} WID \cup \{ \langle\langle A\text{¤}A,I \rangle,\beta \rangle \}, & \text{if DoID}(\alpha,\beta) \text{ involves checking of a weak feature} \\ WID, & \text{otherwise} \end{cases}$

As already indicated, DoID differs from DoIC, in that it keeps (reduction of) checking resources apart from the building of MC/MD-graphs, the former being confined to "terminals," the latter to "pointers." ID, of course is the *immediate dominance* relation.

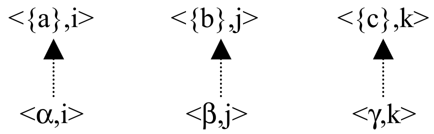
---

[87] Cf. sections 2.6.3, 2.7, and 3.3.2 for comments on the status of labels wrt copying, identity, and extensionality. Stabler (1996, 1998), building on binary branching constituent structure trees, reduces the non-terminal vocabulary of his minimalist system to $<$ and $>$. These labels indicate which daughter of a given node the checking resources can be found on.

[88] I use notation A[x/y] for saying that "x is substituted for y in A." As within the DoIC-system, I assume that the arguments of DoID(α,β) are ordered such that the first one "projects."

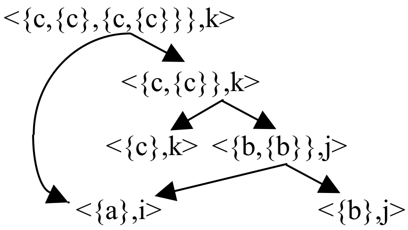
Let me illustrate the two-layered process giving rise to MC/MD-graphs in the DoID-system. First, we create preterminals from terminals via Select. This is sketched in (137).

(137)



The preterminals then figure as pointers, providing the input to DoID in the building of an MC/MD-graph. One possible result is illustrated in (138), for which the crucial part of the derivational transitions is given in (139).

(138)



(139) a. $\mathbb{N}_P$ = { <{a},i>, <{b},j>, <{c},k> }
ID = ∅
b. DoID(<{b},j>,<{a},i>)
c. $\mathbb{N}_P$' = { <{c},k> }
ID' = { <<{b,{b}},j>,<{b},j>>, <<{b,{b}},j>,<{a},i>> }
d. DoID(<{c},k>,<{b,{b}},j>)
e. $\mathbb{N}_P$'' = ∅
ID'' = { <<{b,{b}},j>,<{b},j>>, <<{b,{b}},j>,<{a},i>>,
<<{c,{c}},k>,<{c},k>>, <<{c,{c}},k>,<{b,{b}},j>> }
f. DoID(<{c,{c}},k>,<{a},i>)
g. $\mathbb{N}_P$''' = ∅
ID''' = { <<{b,{b}},j>,<{b},j>>, <<{b,{b}},j>,<{a},i>>,
<<{c,{c}},k>,<{c},k>>, <<{c,{c}},k>,<{b,{b}},j>>,
<<{c,{c},{c,{c}}},k>,<{c,{c}},k>>,
<<{c,{c},{c,{c}}},k>,<{a},i>> }

Analogous to the DoIC-system, DoID brings about multidominance by applying to the same pointer twice. In (138)/(139), this concerns <{a},i>, which is dominated by both <{b,{b}},j> and <{c,{c},{c,{c}}},k>, the appropriate pairs being members of ID.

Note here, that whatever our ultimate stance on extensionality may be (cf. 3.3.2), in the DoID-system its role is rendered harmless. This is due to the fact that checking resources are not involved in the definition of multidominance. I count this as consolation to the skeptic.[89]

Now, given that $ID = IC^{-1}$, the formal relation of the DoIC-system to rooted MC/MD-graphs carries over to the DoID-system. For discussion, I thus refer the reader back to section 3.3.1. What remains to be done here is to transpose the (counterpart of the) Single Root Condition, the Ancestor Condition, and the principle for SYN>PHON-translation. Their adapted versions are given in (140), (141), and (142), respectively.

(140) $SYN = \Sigma\ (= <<\mathbb{N}_T, \mathbb{N}_P>,ID,WID>)$ is translatable into PHON/SEM, only if
$(\exists\alpha \in dom(ID))(\forall\beta \in \mathbb{N}_P \cup dom(ID) \cup ran(ID))[\ <\alpha,\beta> \in ID^*\ ]$

(141) Ancestor Condition on DoID
$\forall\alpha,\beta$, DoID(α,β) can only apply if
a. $\neg\exists\gamma[\ <\gamma,\alpha> \in ID\ ]$, and
b. $\forall\delta[\ <\delta,\beta> \in ID \rightarrow <\alpha,\delta> \in ID^*\ ]$

(142) $\forall\alpha,\beta,\ <\alpha,\beta> \in ID^{\varphi}$ iff
a. $<\alpha,\beta> \in ID\text{-}WID$, and
b. $\neg\exists\gamma[\ <\gamma,\beta> \in ID\text{-}WID \wedge <\gamma,\alpha> \in (ID\text{-}WID)^+\ ]$

I suggest that the addition of indexation together with the separation of terminals from pointers immunizes the DoID-system against the potentially improper appeal to extensionality of its DoIC counterpart, as discussed in 3.3.2. From that counterpart, the DoID-system inherits the power of closely simulating minimalist syntax as developed in Chomsky (1995a) while at the same time meeting hypotheses 3 - 6 (cf. section 2).

---

[89] For the hyperskeptic, a representation of MC/MD-graphs can be devised that lists every node exactly once. Thus, given that indexation is available, we can supply each pointer with two sets of indices, the first one containing indices shared with an immediate ancestor ("up"), the second one containing indices shared with an immediate successor ("down"). A graph like (138) could thus be coded as a set of nodes. This set can be derived bottom-up by applications of a counterpart of DoID as follows. (For the sake of brevity, I ignore projection relations and use *d*, *e*, and *f* instead.)

(i) a. $\{<a,\varnothing,\varnothing>, <b,\varnothing,\varnothing>, <c,\varnothing,\varnothing>\}$ b. DoID(b,a)
c. $\{<a,\{1\},\varnothing>, <b,\{1\},\varnothing>, <c,\varnothing,\varnothing>, <d,\varnothing,\{1\}>\}$ d. DoID(c,d)
e. $\{<a,\{1\},\varnothing>, <b,\{1\},\varnothing>, <c,\{2\},\varnothing>, <d,\{2\},\{1\}>, <e,\varnothing,\{2\}>\}$ f. DoID(e,a)
g. $\{<a,\{1,3\},\varnothing>, <b,\{1\},\varnothing>, <c,\{2\},\varnothing>, <d,\{2\},\{1\}>, <e,\{3\},\{2\}>, <f,\varnothing,\{3\}>\}$

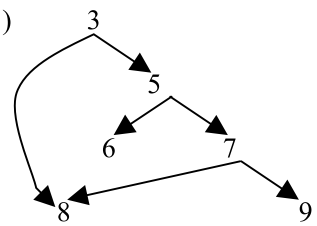
3.4 Some Elaborations of MC/MD-Syntax

In this section, I discuss what is needed to supplement MC/MD-syntax with C-Command (3.4.1), X°-movement (3.4.2), and linear precedence (3.4.3). In the case of X°-movement, the discussion is inconclusive leaving a detailed study of the options sketched for further research.

3.4.1 C-Command[90]

It may appear slightly surprising that the notion of C-Command has not yet been defined explicitly in MC/MD-syntax, given its all-pervasive presence in minimalist syntax and its predecessors. However, one of its central uses, namely, the C-Command Condition on movement, has been restated in terms of the Ancestor Condition on DoIC/DoID (cf. sections 3.2, 3.3.1, and 3.3.3). Also, given that I won't have anything to say about locality and scope, C-Command will remain a side-issue. Here, I'll only explore some options.
Consider once again a graph like (143).

(143)



Clearly, given multidominance, there are at least two ways of defining C-Command in terms of nodes. Call these "Strong" and "Weak C-Command," as given in (144) and (145), respectively.

(144) Strong C-Command (SCC)
α *strongly c-commands* β, iff
a. $\alpha \neq \beta$, and
b. $<\alpha,\beta> \notin ID^+$, and
c. $\forall\gamma[<\gamma,\alpha> \in ID \rightarrow <\gamma,\beta> \in ID^+]$

90 I will have nothing to say about derivational versions of C-Command, as discussed in Epstein (1999) and Epstein et al. (1998).

(145) Weak C-Command (WCC)
α *weakly c-commands* β, iff
a. $\alpha \neq \beta$, and
b. $\langle\alpha,\beta\rangle \notin ID^{+}$, and
c. $\exists\gamma[\ \langle\gamma,\alpha\rangle \in ID \wedge \langle\gamma,\beta\rangle \in ID^{+}\ ]$

Let me also introduce two auxiliary notions, namely, the set of nodes a given node (strongly/weakly) c-commands and the set of nodes a given node is (strongly/weakly) c-commanded by. Call these the “(strong/weak) command domain“ and the “(strong/ weak) commandee domain,“ respectively.

(146) (Strong/Weak) Command Domain
$(S/W)CC(x) =_{def} \{\ y \mid \langle x,y\rangle \in (S/W)CC\ \}$

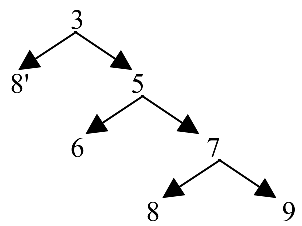
(147) (Strong/Weak) Commandee Domain
$(S/W)CC^{-1}(x) =_{def} \{\ y \mid \langle x,y\rangle \in (S/W)CC^{-1}\ \}$

Applying this to the multidominated node 8 in (143), we arrive at the following result.

(148) a. $SCC(8) = \{9\}$ c. $WCC(8) = \{5,6,7,9\}$
b. $SCC^{-1}(8) = \{6,9\}$ d. $WCC^{-1}(8) = \{6,9\}$

Compare this to standard minimalist syntax, where dependency formation leads to creation of an additional node. Instead of (143), we get a representation like (149), in which 8' is a “copy“ of 8, the two nodes forming a chain.[91]

(149)



Of course, given the Single Mother Condition on constituent structure trees, SCC and WCC are equivalent here. I'll therefore speak of just CC when discussing standard structures. (150) then gives the resulting domains for nodes 8 and 8' of (149).

(150) a. $CC(8) = \{9\}$ c. $CC(8') = \{5,6,7,8,9\}$
b. $CC^{-1}(8) = \{6,8', 9\}$ d. $CC^{-1}(8') = \{5\}$

[91] The following discussion may count as further evidence that strict identity of copies is not intended in minimalist syntax. Such a construal would necessitate an adjustment of C-Command along the lines proposed here. I'm not aware of any attempt within standard minimalism to provide such an adjustment.

Reflexivity aside, the two approaches have the following characteristics. From a dynamic perspective, standard movement adds a node with an extended command domain, i.e. $CC(8) \subset CC(8')$. The commandee domain of the added node acquires one element, namely, its sister, while everything dominated by its sister is removed. Thus, in (149), $CC^{-1}(8') = (CC^{-1}(8) - \downarrow 5) \cup \{5\}$.[92] In MC/MD-syntax, extending the command domain of 8 in (143) is a straightforward matter of switching from SCC to WCC. However, under the current perspective, dependency formation (i.e. creation of links) does not alter the commandee domain of nodes. In fact, $SCC^{-1}(8) = WCC^{-1}(8)$. This is a substantial enough shortcoming, so I will return to it shortly.

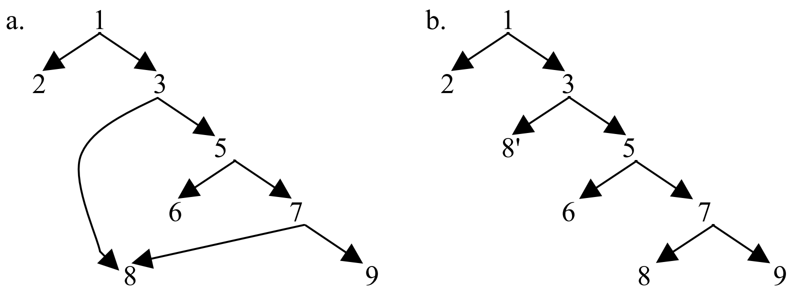
Reflexivity is ruled out by (144a)/(145a). Thus, 8 does not c-command itself in MC/MD-syntax, i.e. $8 \notin SCC/WCC(8)$ and $8 \notin SCC^{-1}/WCC^{-1}(8)$. This ban on reflexive C-Command carries over *mutatis mutandis* into the standard system.[93] However, 8' is able to c-command its copy, 8, given their non-identity. It follows that $8 \in CC(8')$ and $8' \in CC^{-1}(8)$. Nevertheless, this seems to be a marginal issue, given the fact that the C-Command Condition on movement/chains is reformulated in MC/MD-syntax in terms of the Ancestor Condition on DoIC/DoID. No direct appeal to C-Command it required.

Let me return to the different commandee domains of 8 in (143) vs. 8' in (149). The following example shows a potentially unwelcome consequence of such a difference. Thus, consider the (simplified) definition of "closeness" in (151).[94]

(151) α is closer to γ than β is, iff
a. $\{\alpha,\beta\} \subseteq CC(\gamma)$, and
b. $\beta \in CC(\alpha) \wedge \alpha \notin CC(\beta)$

(152), which adds some structure to (143) and (149), brings out the essential point.

(152)



We want at least in principle to be able to say wrt (152a) that 8 is closer to 2 than 6 is. It is easy to verify that the counterpart to that statement wrt (152b), i.e. that 8' is closer to 2 than 6 is, holds. The problem with (152a) is that, in violation of (151b), $6 \in WCC(8)$

---

[92] See section 3.2 for this notation as well as a discussion of Proper Upper Cones and C-Command.

[93] Barker&Pullum (1990) argue on formal grounds that the ban on reflexivity should not be part of the definition of C-Command. Rather its effects should be derived by independent means.

[94] For comprehensive discussion, see Chomsky (1995a, p.299, p.355) and Ferguson (1996).

and $8 \in$ WCC(6).[95] In order to get the desired result, we have to shift our attention from 8 and 6 themselves to their immediate ancestors. Obviously, we can define the required set as follows.

(153) Immediate Ancestors
$IA(x) =_{def} \{ y \mid \langle y,x \rangle \in ID \}$

Now, 8 and 6 have different sets of immediate ancestors, specified in (154).

(154) a. $IA(8) = \{3,7\}$
b. $IA(6) = \{5\}$

We can then utilize the asymmetry between these sets, i.e. $\langle 3,5 \rangle \in ID^{+}$ and $\langle 5,3 \rangle \notin ID^{+}$, to derive revised notions of C-Command that solve our problem. The required definitions are given in (155) and (156).[96]

(155) Strong C-Command (SCC) [revised]
α *strongly c-commands* β, iff
a. $\alpha \neq \beta$, and
b. $\langle \alpha,\beta \rangle \notin ID^{+}$, and
c. $(\forall\gamma \in IA(\alpha))(\forall\delta \in IA(\beta))[\ \langle \gamma,\delta \rangle \in ID^{*}\ ]$

(156) Weak C-Command (WCC) [revised]
α *weakly c-commands* β, iff
a. $\alpha \neq \beta$, and
b. $\langle \alpha,\beta \rangle \notin ID^{+}$, and
c. $(\exists\gamma \in IA(\alpha))(\forall\delta \in IA(\beta))[\ \langle \gamma,\delta \rangle \in ID^{*}\ ]$

If we now substitute the revised WCC for CC in the definition of closeness, we get the desired outcome, i.e. $6 \in$ WCC(8) and $8 \notin$ WCC(6).

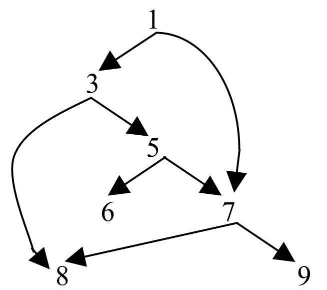
An additional subtlety comes to the fore if we consider remnant movement configurations. Recall that these do not pose any problem for SYN>PHON-translation in MC/MD-syntax, as shown in section 3.3.1. As far as C-Command is concerned, however, some extra fine-tuning is unavoidable. The graph in (157) shows the configuration in question.

---

95 For MC/MD-graphs, WCC has to be substituted for CC in the definition of closeness, of course.

96 Going from ID^{+} to ID^{*} is necessary for allowing mutually c-commanding sister nodes.

(157)



Derivationally speaking, after 8 has been "extracted" from ⇃7, the latter is "raised." Again, we want to at least in principle have the option of saying that 7 in (157) c-commands 8. However, clause (156b) of WCC precludes this, given that $<7,8> \in ID^{+}$.[97] An additional adjustment is clearly called for. This time, it is the notion of dominance that must be refined. I suggest that "Strict Dominance," as defined in (158), serves that purpose best.

(158) Strict Dominance (SD)
$\alpha SD\beta \Leftrightarrow_{def} \forall\gamma[\ \gamma \in IA(\beta) \rightarrow <\alpha,\gamma> \in ID^*\ ]$

On the basis of (158), we can formulate our final revision of SCC and WCC, provided in (159) and (160), respectively.

(159) Strong C-Command (SCC) [final revision]
α *strongly c-commands* β, iff
a. $\alpha \neq \beta$, and
b. $<\alpha,\beta> \notin SD$, and
c. $(\forall\gamma \in IA(\alpha))(\forall\delta \in IA(\beta))[\ <\gamma,\delta> \in ID^*\ ]$

(160) Weak C-Command (WCC) [final revision]
α *weakly c-commands* β, iff
a. $\alpha \neq \beta$, and
b. $<\alpha,\beta> \notin SD$, and
c. $(\exists\gamma \in IA(\alpha))(\forall\delta \in IA(\beta))[\ <\gamma,\delta> \in ID^*\ ]$

Reconsider remnant movement in the light of these modifications. Clearly, 7 does not strictly dominate 8 in (157), given that $3 \in IA(8)$ and $<7,3> \notin ID^*$. Thus, the pair <7,8> qualifies for WCC, as intended.[98]

---

[97] See again Barker&Pullum (1990), for formal reasons not to include any such dominance requirement in the definition of C-Command.

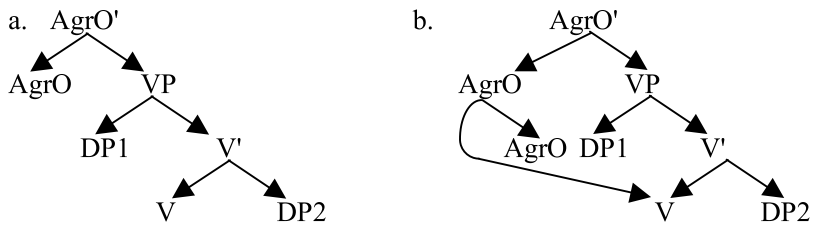
[98] Note that a more procedural perspective on C-Command and closeness does not seem to privilege standard minimalist syntax over MC/MD-syntax. Thus, assume an operation like Attract (cf. Chomsky 1995a, chapter 4) includes a top-down search procedure for identifying compatible features. Clearly, immediate dominance relationships involving moved constituents in constituent

Given the formal results in Kracht (2001) concerning multidominance and copy chains, we can be confident that further refinements along the same lines are possible if required. Clearly, the area of covert movement would need some extra consideration, which, however, I will not go into here.[99]

3.4.2 X°-Movement

It has repeatedly been noted earlier that standard X°-movement has to be treated as an exception to notions of strict cyclicity like the “extension condition“ on GT/Merge (cf. 2.5.2, 2.6.3). MC/MD-syntax brings out this property very clearly as well. Thus, consider the transition from (161a) to (161b).[100]

(161) a. b.



---

tree-like structures will have exact counterparts involving multidominance in MC/MD-graphs. Consequently, construing closeness in terms of steps on an “ID-path“ does not distinguish the two approaches.

[99] Most distinctions in that domain will have to rely on whether principles are formulated on the basis of ID or ID-WID. Thus, consider the following set of data, discussed by Brody (1995, p.133).

(i) a. John$_i$ wondered [$_{CP}$ [$_{DP1}$ which pictures of himself$_i$]$_k$ [$_{IP}$ Mary saw t$_k$]]

b. * John$_i$ wondered [$_{CP}$ [when] [$_{IP}$ Mary saw [$_{DP1}$ which pictures of himself$_i$]]]

c. John wondered [$_{CP}$ [$_{DP1}$ which pictures of himself]$_k$ [when] [$_{IP}$ Mary saw t$_k$]]

(ia) illustrates overt *WH*-movement of DP1, which is compatible with anaphor binding by the matrix subject *John*. (ib) indicates that no such binding is possible if DP1 stays in situ. This is problematic in that (ic), the LF representation of (ib) resulting from covert *WH*-movement of DP1, is supposed to be the relevant input to binding-theoretic computations, the prediction being that (ib) should be well-formed. The problem does not arise in Chomsky (1995a, chapter 4), where movement after Spell-Out is restricted to feature movement. Thus, only FFORM of DP1, located in D° as far as I can see, would move to COMP. The anaphor stays put and no local binding between *John* and *himself* can be established. MC/MD-syntax does not employ feature movement. Yet, the crucial difference between (ia) and (ib) is that <CP,DP1> is a member of WID in the latter but not in the former. A proper formulation of binding theory can be sensitive to such a difference allowing MC/MD-syntax to have it both ways. Constitutents can be linked into more than one syntactic context, while decisions on what that implies are dependent on the module of grammar accessing this abstract representation. For further consideration of the above field of data see among others Brody (1995, chapter 5), Engdahl (1986), Heycock (1995), Lebeaux (1991), and Reinhart (1995).

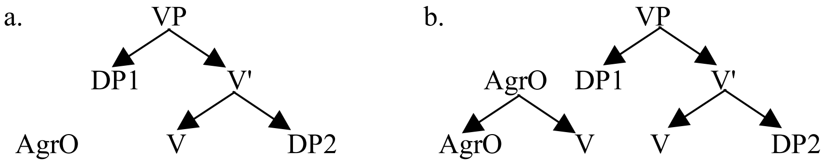
[100] I sidestep the question of dynamic/relative bar-level determination here. Thus, AgrO' rather than AgrOP is the root in (161).

This transition is in conflict with the "Strictest Cycle Condition" implied by the DoIC- and DoID-system. In fact, the Ancestor Condition prohibits DoID(AgrO,V) from applying at stage (161a), given that $\langle AgrO',AgrO\rangle \in ID$. This is independent of the question whether the output structure should involve X°-adjunction or which notion of C-Command would allow V to c-command its trace/copy in the standard alternative to (161b).[101]

In the following, I will briefly sketch four approaches toward reconciling X°-movement with MC/MD-syntax. The first two will among other things require some relativization of the Ancestor Condition and would thus call for a more drastic modification than the other two. All four approaches will have to be supplemented with more specific assumptions about checking and the internal make-up of lexical items. Such information will be considered only very cursorily.

Bobaljik (1995a) and Bobaljik&Brown (1997) advocate a "sideward movement" approach to X°-movement. Accordingly, their counterpart of (161b) involving copies of V is not derived from (161a) directly but via (162a) and (162b).

(162)



Note that the latter transition would not violate the Strictest Cycle Condition. In that respect, the Ancestor Condition, which rules them out, is more restrictive than imposing the Strictest Cycle Condition directly (cf. section 3.3.1). It would, thus, be an option to weaken the Ancestor Condition on DoID relative to what kind of feature-checking is involved. This would have to be based on the distinction between X°- vs. XP-features (a.k.a. V- vs. N-features), the former checked via X°-movement, the latter via XP-movement (cf. section 2.4). The required "Relativized Ancestor Condition on DoID" is given in (163).

(163) Relativized Ancestor Condition on DoID[1]
$\forall\alpha,\beta$, DoID(α,β) can only apply if
a. $\neg\exists\gamma[\ \langle\gamma,\alpha\rangle \in ID\ ]$, and,
unless it involves checking of X°-features,
b. $\forall\delta[\ \langle\delta,\beta\rangle \in ID \rightarrow \langle\alpha,\delta\rangle \in ID^*\ ]$

In addition to this weakening of the Ancestor Condition, it has to be made sure that X°-dependencies observe locality, in particular the Head Movement Constraint (HMC). Bobaljik&Brown (1997, p.351f) suggest that this is guaranteed by a "Chain Condition," incorporating C-Command and closeness. This condition is checked on every application of Merge. Thus the transition from (162b) to the counterpart of (161b)

[101] See Sternefeld (1991) for technical discussion of X°-movement in pre-minimalist frameworks. For a formalized representational approach, see Cornell (1999).

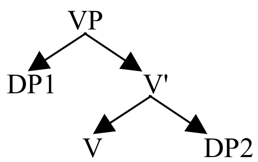
would be well-formed, given that V, if construed as adjoined to AgrO, locally c-commands its trace under standard assumptions.

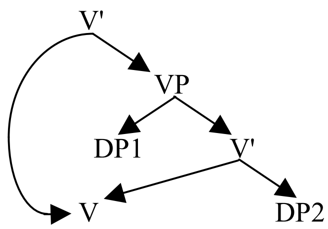
This proposal has two unattractive features as far as MC/MD-syntax is concerned. First, I would prefer adjunction to be avoided.[102] Secondly, given that the C-Command Condition on movement has been replaced by the Ancestor Condition on DoID, it would be counterintuitive to add Bobaljik&Brown's (1997) rather global Chain Condition to just capture the HMC. A somewhat *ad hoc* and much more limited principle such as (164) may be sufficient for avoiding both these unattractive features.

(164) Locality of X°-Dependencies
DoID(α,β) is blocked if
$\exists\gamma,\delta[\ \gamma \approx \delta \wedge \langle\alpha,\gamma\rangle \in ID^{+} \wedge \langle\beta,\gamma\rangle \in ID^{+} \wedge \langle\beta,\delta\rangle \in ID^{+} \wedge$
$\gamma \in WCC(\delta) \wedge \delta \notin WCC(\gamma)\ ]$

(164) says that a sidewardly moved element γ cannot be rejoined to the subgraph it originates from if that subgraph contains an item δ "comparable" to γ, (γ ≈ δ), that asymmetrically c-commands γ. Of course, it would have to be made more precise what exactly is meant by comparability. This depends among other things on a proper notion of complex heads.[103]

The approach by Ackema et al. (1993) allows for lexical heads like V° to project more than once. In that respect, they could be considered to provide their own landing sites, "Münchhausen-style." Consider the MC/MD-graphs in (165).

(165) a.


b.


For the MC/MD-system developed in section 3.3., this would be attractive insofar as "sideward movement" is avoided. However, the projecting argument of DoID is not a root at the stage where DoID should apply. Thus, again, the Ancestor Condition is violated. Another relativization, along the lines of (166), would, of course, be possible.

---

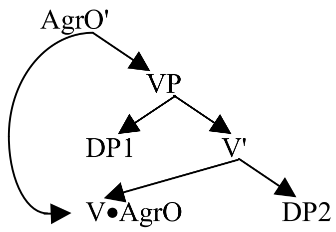
[102] This is also envisaged in Chomsky (1998, p.127), where it is stated that "[. . .] optimal design should [. . .] perhaps remove from the core syntax such operations as adjunction of categories (XP-adjunction and head-adjunction)." For technical studies of adjunction, see Kolb (1997b) and Kracht (1999).

[103] See e.g. Chomsky (1995a), Bobaljik (1995b), and Zwart (1997) for discussion.

(166) Relativized Ancestor Condition on DoID[2]
$\forall\alpha,\beta$, unless α is a preterminal, and β a projection of α, and
$\neg\exists\gamma[\ \langle\gamma,\beta\rangle \in \text{ID}\ ]$,
DoID(α,β) can only apply if
a. $\neg\exists\gamma[\ \langle\gamma,\alpha\rangle \in \text{ID}\ ]$, and,
b. $\forall\delta[\ \langle\delta,\beta\rangle \in \text{ID} \rightarrow \langle\alpha,\delta\rangle \in \text{ID}^*\ ]$

Yet, having the same preterminal project twice will, given the interpretation of extensionality underlying the DoID-system, lead to a ternary branching V' rather than a new projection. Thus, DoID(V,VP) would add <V',V>, redundantly, and <V',VP> to ID, yielding ID' = { <V',V>, <V',DP2>, <V',VP>, <VP,V'>, <VP,DP1> }. In fact, this would even violate the Acyclicity of (rooted) MC/MD-graphs (cf. secton 3.2). An adequate modification would have to make richer assumptions about lexical items heading an "extended projection" in the sense of Grimshaw (1991). Such items should perhaps be inserted as ordered compounds like V•AgrO, where each part successively projects. Under such a theory, (167) should be derived instead of (165b).

(167)

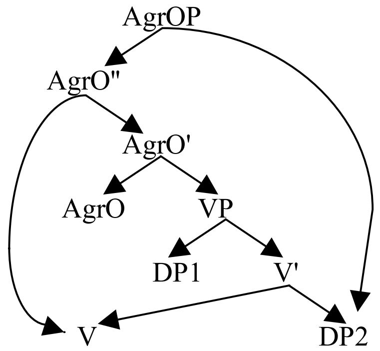


Making the proper adjustments to the MC/MD-system, however, is beyond the scope of the current study.

A third approach to X°-movement can be built on the notion of "multiple specifiers," developed in Koizumi (1995).[104] This has in fact been suggested by Toyoshima (1997) and Fukui&Takano (1998). Here, verb movement could target the "inner" specifier of AgrO while objects target the "outer" one, as illustrated in (168).[105]

[104] Cf. Chomsky (1995a, section 4.10) and Richards (1999).
[105] Recall the MC/MD-graphs are not linearly ordered.

(168)



This solution, it is easy to verify, does not require any adjustments of the DoID-system as introduced in section 3.3. AgrO'' is introduced by an application of DoID(AgrO',V). Of course, what has to be reglemented is the hierarchical order of specifiers. Fukui&Takano (1998) consider such an order a matter of parameterization, ultimately encoded in the functional heads of each language. A principled way of forcing X°-movement into the specifier closest to a functional head may be based on (a generalization of) Bobaljik's (1995b) notion of "morphological merger," which itself is built on adjacency. Some such notion seems to be required in any case for incorporation and cliticization.

Alternatively, a condition can be put on DoID(α,β) such that X°-feature checking between *th*(α) and *th*(β) is blocked as soon as α does not immediately dominate the preterminal of its projection. Take "I(x)" to denote the index of x.[106]

(169) Condition on X°-feature Checking
DoID(α,β) cannot involve checking of X°-features if
$\exists\gamma,\delta[\ \langle\alpha,\gamma\rangle \in \text{ID} \wedge \langle\gamma,\delta\rangle \in \text{ID} \wedge \text{I}(\alpha) = \text{I}(\gamma) = \text{I}(\delta)\ ]$

This condition forces X°-movement into the "innermost specifier" of the target.

On the other hand, the ordering of specifiers may be avoidable under a theory of late lexical insertion, as proposed in Jackendoff (1997). German V2, for example, which is standardly taken to involve XP-movement into Spec,CP and X°-adjunction to C°, could be reconceptualized as the spell-out of a C°, one of whose specifiers contains a "strongly linked" V°, i.e. which has checked a strong X°-feature against V°. V-final counterparts could then be the spell-out of a C°, realized for example as *dass*, one of whose specifiers contains a "weakly linked" V°, i.e. which has checked a weak X°-feature against V°. This would basically be equivalent to the (lexical) stipulation that the former C° does, while the latter does not allow V° to adjoin to it. It is hard to see how theories of V2 can do without any lexico-morphological assumption along similar lines. Under this perspective, an X° in a specifier position would never itself be

[106] This condition would leave the option of X°-feature checking under sisterhood, i.e. by application of "Merge" in the terminology of minimalist syntax. It is actually unclear, whether standard minimalist syntax can do without such a condition.

phonologically realized. Rather it would license a special spell-out on the head providing that specifier as a target.

Finally, a proposal related to the previous one would not allow X°-movement at all. Again, it is clear why such an approach is compatible with the MC/MD-system. Of course, lexical assumptions have to be made about the actual spell-out position each lexical head obtains in its extended projection. Such an approach would have to operate checking between heads at the point where selectional requirements are satisfied as well. Stabler (1996, 1998) argues, that the Merge operation of syntax is the right mechanism to do this. Thus, DoID(Agr,VP), yielding an MC/MD-graph like (161a), would check both X°- and selectional features of AgrO° and V°.

This is the first point where my (conditional) hypothesis 2 from section 2.4 would come into play.[107]

H2: Under certain conditions, (part of) the theta criterion has to be implemented as a syntactic well-formedness constraint

The following strong condition on DoID would then provide the background for this implementation.

(170) Checking Condition on DoID
Every application of DoID involves feature checking

The checking of selectional features must be one way of satisfying (170).

I will have to leave things in this somewhat sketchy and inconclusive state. Note, however, that X°-movement is currently undergoing rather radical scrutiny from widely diverging perspectives. One line of research attempts to locate its properties in the PF-component.[108]

> "There are some reasons to suspect that a substantial core of head-raising processes, excluding incorporation in the sense of Baker (1988), may fall within the phonological component" (Chomsky 2001, p.37).

The opposite line is an attempt to fully reduce X°-movement to XP-(remnant) movement. This is pursued in Koopman&Szabolcsi (2000). It may therefore be premature to rule in or rule out any of the above options for treating X°-movement in MC/MD-syntax.

3.4.3 Linear Precedence

Recall from section 2.6 that the minimalist approach to linear precedence consists in projecting precedence from asymmetric C-Command on the basis of Kayne's (1994)

[107] See section 3.5, for another appeal to H2.
[108] For a critique of this position, see Zwart (2001).

LCA. Given the availability of C-Command in MC/MD-syntax (cf. section 3.4.1), we could follow this line as well. Note also that it is unclear at this point whether linearization should occur at Spell-Out or further removed from syntax on the PF-branch of C_{HL}.[109]

In this section, however, I would like to briefly illustrate how linear precedence can already be partially introduced into the SYN-component of the MC/MD-system. The idea is to impose an ordering on "siblings" in MC/MD-graphs, much as originally proposed by Peters&Ritchie (1981) (cf. section 3.2).[110] Independently, Saito&Fukui (1998) suggest that linearity should be directly captured by (a version of) Merge, which produces not sets but ordered pairs of constituents, order being interpretable as left/right ("early"/"late") distinction. The structures generated by Merge are characterized in (171) (Saito&Fukui 1998, p.455).[111]

(171) $K = \{ \gamma, < \alpha, \beta > \}$, where $\gamma \in \{ \alpha, \beta \}$
a. $\gamma = \alpha$: head-initial, left headed
b. $\gamma = \beta$: head-final, right headed

Combining the two approaches above, we can introduce a "sibling precedence"(SP-) relation into the DoID-system. This will enrich derivational stages with another binary relation on pointers, so we arrive at $<\mathbb{R}$,ID,WID,SP$>$-quadruples. The update of SP resulting from an application of DoID(α,β) is given in (172).

(172)
$$SP' = \begin{cases} SP \cup \{<\alpha,\beta>\}, \text{ if p} \\ SP \cup \{<\beta,\alpha>\}, \text{ if q} \end{cases}$$

It has to be assumed that *p* exhaustively states the factors requiring $\alpha<\beta$, while *q* does so for $\beta<\alpha$. This can be made more concrete as soon as a linguistic theory of linear precedence is added. Thus, application of DoID could be sensitive to either the head parameter, or features encoded in the terminal heads. Alternatively, on applying DoID(α,β) a set of global LP-rules in the sense of Gazdar et al. (1985) could be consulted.[112]

Recall from section 3.2 that projecting a strict partial precedence order from sibling precedence is not in general possible as long as SP is defined on MC/MD-graphs. The projection rules are repeated in (173).

[109] For discussion, see Nunes (1995, 1999) and Uriagereka (1997, 1999). This has a bearing on the syntax-morphology relation (cf. section 3.5).

[110] Stabler (1996, 1998) defines minimalist grammar in terms of constituent structure trees. Thus linear precedence is fully part of syntax proper in his system.

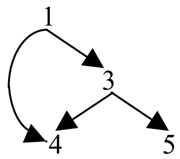
[111] Terminology reveals that it is the head parameter Saito&Fukui (1998) are most interested in.

[112] See also Ojeda (1988) among others.

(173) Precedence
a. $(\forall x,y \in N)[\ \langle x,y\rangle \in SP \rightarrow \langle x,y\rangle \in P\ ]$
b. $(\forall x,y,z,w \in N)[\ (\ \langle x,y\rangle \in P \wedge z \in \downarrow x \wedge w \in \downarrow y\ ) \rightarrow \langle z,w\rangle \in P\ ]$

Consider again a simple case, such as (174).

(174)



Whichever pair of $\langle 3,4\rangle$ or $\langle 4,3\rangle$ is in SP, $\langle 4,4\rangle \in P$, given (173). In order to avoid this, we have to resolve the ambiguous constituency for multidominated pointers. Recall that this task has been accomplished by ID^{φ} already. We can therefore define SP^{φ} with reference to that reduction of ID, as given in (175).

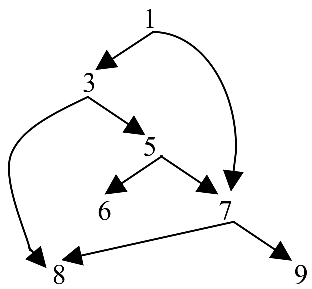
(175) For every $\langle\alpha,\beta\rangle \in SP$, $\langle\alpha,\beta\rangle \in SP^{\varphi}$ iff
$\exists\gamma[\ \langle\gamma,\alpha\rangle \in ID^{\varphi} \wedge \langle\gamma,\beta\rangle \in ID^{\varphi}\ ]$

(175) says that in SP^{φ} we only register precedence information on siblings that continue to be siblings in ID^{φ}. A strict partial precedence order now results if we substitute SP^{φ} for SP in (173) and interpret $\downarrow^{\varphi}$ in terms of ID^{φ}.

(176) φ-Precedence
a. $(\forall x,y \in N)[\ \langle x,y\rangle \in SP^{\varphi} \rightarrow \langle x,y\rangle \in P\ ]$
b. $(\forall x,y,z,w \in N)[\ (\ \langle x,y\rangle \in P \wedge z \in \downarrow^{\varphi} x \wedge w \in \downarrow^{\varphi} y\ ) \rightarrow \langle z,w\rangle \in P\ ]$

Let me illustrate this with the case of remnant movement discussed in section 3.4.1, repeated below as (177).

(177)



Assume the SP-relation as given in (178).

(178) $SP = \{\ \langle 7,3\rangle, \langle 8,5\rangle, \langle 6,7\rangle, \langle 8,9\rangle\ \}$

Also assume that WID is empty, i.e. no weak features have been checked. Thus ID^{φ} will register the “higher“ links of multidominated nodes. This is given in (179).

(179) ID^{φ} = { <1,3>, <1,7>, <3,8>, <3,5>, <5,6>, <7,9> }

On the basis of (178) and (179) we derive SP^{φ} in (180).

(180) SP^{φ} = { <7,3>, <8,5> }

Finally, from (180) we can project P, given in (181), in terms of (176).

(181) P = { <7,3>, <8,5>, <7,8>, <7,5>, <7,6>, <9,3>, <9,8>, <9,5>, <9,6>, <8,6> }

This, indeed, is the desired outcome. It shows that, following Peters&Ritchie (1981), precedence information can be handled by MC/MD-syntax, making “sibling precedence“ available throughout and standard precedence projectable for the PF-component.

3.5 Some Objections to MC/MD-Syntax

Let me finally address some objections to MC/MD-syntax. First, it must be asked how “freezing-effects“ of overt movement on covert extraction can be captured under strictest cyclicity. Secondly, it is unclear how to deal with PF-phenomena that seem to involve the “spelling-out“ of multiple “copies“ or “traces“ of a single constituent. Finally, it has to be shown how differences between Merge and Move related to a Θ-theoretic asymmetry between the base- and the transformational syntactic component are handled. I believe that only a subset of the second kind of objections requires modifications of a less than tolerably *ad hoc* nature. So, this is where further in depth work is called for. Let me consider each objection in turn.

Collins (1997) argues for the necessity of counter-cyclic movement and thus against “single pass“ approaches in general[113] and MC/MD-syntax, built to satisfy hypothesis 3, in particular. Consider the *prima facie* problematic case in (182) (Collins 1997, p.89).

(182) * [John to be nice], I consider

One way of ruling out (182) would be to assume that the ECM-subject *John* cannot check Case features in Spec,AgrOP of the matrix clause since that would require lowering.[114] Indeed, lowering is not an option for any of the frameworks under

---

[113] See section 3.3.1 and references cited there.

[114] Under a copying approach, it has to be prevented that the “copy“ of *John* inside the copy of *John to be nice* undergoes LF-movement. Chomsky (1995a, p.304) rules this out by the principle in (i).

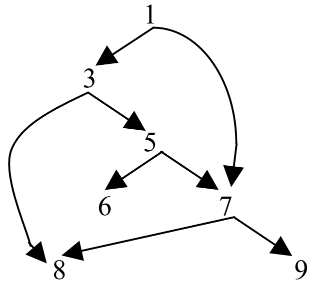
(i) Trace is immobile

discussion here. Collins further suggests the following general principle to be responsible for the ill-formedness of (182) (Collins 1997, p.89).

(183) Suppose α contains a constituent β and α undergoes overt movement. Then β may not undergo covert movement to a position outside of α.

Note that this is a variant of the "freezing" principle going back to Ross (1986), according to which certain cases of overt movement lead to the opacity of the displaced constituent. Now, assuming (183) to be correct, it could be recast in MC/MD-syntax as a constraint on "pseudo-remnant movement." Thus, abstractly, (182) conforms to our earlier picture of remnant movement, repeated for convenience in (184).

(184)



John would correspond to 8, and *John to be nice* to ⇃7. Covert movement would be captured by entering <3,8> into both ID and WID after DoID(5,8) has applied. The required constraint simulating (183) can then be formulated as in (185).

(185) Ban on Pseudo-Remnant Movement
DoID(α,β) is blocked if
$\exists\gamma,\delta[\ <\gamma,\delta> \in \text{WID} \wedge <\gamma,\beta> \in \text{ID}^+ \wedge <\beta,\delta> \in \text{ID}^+\ ]$

Thus, a constituent ⇃β is "frozen in place," if it contains a constituent ⇃δ weakly linked to the outside of ⇃β. In this sense, ⇃β is just a *pseudo*-remnant. As we have already seen, standard remnant movement, where $<\gamma,\delta> \notin \text{WID}$, must be allowed.[115]

However, it is unclear to what extent (183) is a general principle in the first place. Thus, consider the examples in (186), (186a) from Swedish, (186b)-(186d) from German.[116]

See section 2.7.2 for more fundamental further arguments against such an approach.

[115] Cf. section 3.3.1, as well as Müller (1998) and papers in Alexiadou et al. (eds.)(forthcoming).

[116] (186b) goes back to Haider (1990).

(186) a. $[_{CP}$ $[_{VP}$ Köper bröd $]_i$ [gör $[_{IP}$ jag inte t_i]]]
buy bread do I not
b. $[_{CP}$ [Aussenseiter gewonnen $]_i$ [haben $[_{IP}$ diesen Preis noch nie t_i]]]
outsiders won have this prize yet never
"Outsiders haven't ever won this prize so far."
c. Wer hat [den Peter wann zu besuchen] versprochen
who has the Peter when to visit promised
"Who promised to visit Peter when"
d. Wer hat versprochen [den Peter wann zu besuchen]

(186a) and (186b) show VP-fronting where VP contains an argument, a direct object in (186a) and a subject in (186b). This argument has to check its Case inside the stranded IP. As for (186c) and (186d), it is clear that if covert WH-movement of *wann* takes place, it has to be able to proceed from an overtly moved constituent, in violation of principle (183). This is due to the fact that either (186c) is derived from (186d), or vice versa.[117]

I conclude that MC/MD-syntax is not affected by this first objection, given that (i) its putative technical problem can be solved in a way that preserves strictest cyclicity, and (ii) it is built on an empirical generalization the scope of which must be considered fairly controversial.

Let us next turn to some issues related to SYN>PHON-translation. The reduction method assumed to yield ID^{φ} eliminates all links but one, for each multidominated constituent. This would seem to predict that there is a one-to-one relationship between constituents and positions in an output string. Such a view, however, may be challenged by citing so-called "copy-movement" in German and phenomena like English *wanna*-contraction as well as VP-deletion. I'll address copy-movement first.[118] Consider the sentence in (187).

(187) Wen meinst du wen wir einladen sollen
Who think you who we invite should
"Who do you think we should invite?"

Semantically, (187) constitutes one constituent question, as the translation indicates. Thus, π seems to realize a superfluous copy of *wen*. Standard approaches count (187) as vindication of the copy-theory of movement, allowing, exceptionally, for the spelling-out of more than one copy of a chain. Of course, this approach is not available to me. An alternative, which I owe to Jairo Nunes (p.c.), is to take more seriously the morphological constraints on this construction. Thus, while WH-copying works best with bare WH-pronouns, it is unacceptable with complex phrases. This is shown in (188).[119]

[117] I assume that participle movement is not an option for deriving this alternation. Collins (1997, p.138fn.29) cites LF-anaphor raising as another counterexample to principle (183).
[118] See also Nunes (1995), and papers in Lutz et al. (eds.)(2000).
[119] Some PPs involving bare WH-pronouns seem to give intermediate results.

(188) * Welche Leute meinst du welche Leute wir einladen sollen
Which people think you which people we invite should
“Which people do you think we should invite?“

It may thus be possible to forge a link between WH-copying and the phenomenon of complementizer agreement, present in some variants of German and Dutch.[120] There, agreement features of the finite verb are spelled out twice, once on the finite verb and once on the complementizer. This is shown for Bavarian in (189).

(189) dass-ts ia liab sei-ts
that-2.Pl you nice are-2.Pl
“that you are nice“

From this perspective, copies of WH-pronouns may be maximally richly agreeing complementizers, having checked features against a WH-operator in Spec,CP. I will have to put off working out the details of such a proposal to future research.

Now, recall from section 2.7.1 that so-called “*wanna*-contraction“ has sometimes been argued to provide evidence for the presence of (Case-marked) traces at PF. I repeat the crucial examples in (190).

(190) a. I want to visit Stockholm
b. I wanna visit Stockholm
c. Who$_i$ do you want t$_i$ to visit Stockholm
d.* Who$_i$ do you wanna t$_i$ visit Stockholm
e. I want Mary to visit Stockholm
f. * I wanna Mary visit Stockholm

Assume that there is an “(optional) contraction rule“ like (191) (Chomsky 1981, p.181).

(191) want+to $\rightarrow$ wanna

Then, clearly, the trace in (190c) would be responsible for the unacceptability of (190d), given the configuration in (192).

(192) * want+*t*+to $\rightarrow$ wanna

In section 3.4.3, I suggested that linear precedence is computed from ID$^{\varphi}$. This is a representation where *who* of (190c)/(190d) is no longer “linked“ to the extraction site, say Spec,IP, i.e. <IP,*who*> $\notin$ ID$^{\varphi}$. However, subsequent work has shown that a purely

(i) ? An was meinst du an was wir denken sollen
At what think you at what we think should
“What do you think we should take into account?“

120 Cf. Bayer (1984), Grewendorf (1988), and Zwart (1997).

linear approach to *wanna*-contraction makes unwelcome predictions. This is shown by examples from Postal&Pullum (1982), given in (193).[121]

(193) a. I don't want [[PRO to flagellate oneself in public] to become standard practice in this monastery]
b. * I don't wanna [[PRO flagellate oneself in public] to become standard practice in this monastery]

In fact, subsequent analyses seem to converge on the assumption that contraction must be dealt with at the syntactic level containing some structural notion such as government.[122]

> "We can put forth the strong claim that the government condition not only supplements the visibility account of adjacency processes but actually supplants it [. . .]. [. . .] the claim that a Case-marked trace is visible and can be distinguished from other empty categories in that it is the only one that blocks PF rules finds no support in contraction [. . .]" (Bouchard 1986, p.101).

For MC/MD-syntax this means, that *wanna*-contraction is dealt with in the component of grammar where a full-fledged ID-representation is still available. For (190c)/(190d) the information that <IP,*who*> $\in$ ID is thus available if necessary.

Yet, Jacobson (1982) presents further π-related evidence that "argue[s] against a theory which posits a single multidominational representation for a sentence" (Jacobson 1982, p.188). Her crucial cases involve VP-deletion.[123]
Thus, consider the paradigm in (194).[124]

(194) a. I want TIM to $\varnothing$
b. ?? I WANT Tim to $\varnothing$
c. ?* TIM$_i$, I asked t_i to $\varnothing$
d. ?* Tim$_i$, I ASKED t_i to $\varnothing$

---

[121] In Chomsky (1981), PRO is taken not to be "visible" to the contraction rule in (191), given its lack of Case. As Barss (1995, p.683) rightly points out, this account would have to be revised in the light of more recent developments according to which PRO receives "null Case" (cf. Chomsky [with Howard Lasnik] 1995a, p.119f).

[122] Cf. Bouchard (1986) and Barss (1995). For a more lexicalist approach, doing completely without traces, see Sag&Fodor (1994).

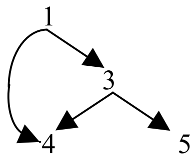
[123] Another argument is built on *tough*-constructions. However, the fact is overlooked that an MC/MD-system is combinable with an "empty operator approach" to these phenomena (cf. Chomsky 1981). Such an approach steers clear of the particular problem pointed out in Jacobson (1982, p.222).

[124] Acceptability judgments for these examples are subtle. I conjecture (194d) from examples independently given in Jacobson (1982).

A number of further subtleties aside, the principle licensing VP-deletion requires stress to fall on the constituent immediately preceding *to*. This is what the contrast in (194a)/(194b) shows. But then the uniform unacceptability of (194c)/(194d) creates a dilemma for MC/MD approaches. Clearly, the level of stress assignment either does or does not contain an edge linking the DP *Tim* into its base position. If it does, (194c) is incorrectly predicted to be fine, given that *Tim* is stressed. If it doesn't, (194d) is incorrectly predicted to be fine, given that *asked* is stressed. Jacobson (1982) concludes that leaving an inherently unstressed "gap" or "trace," in the base position of *Tim*, makes the right predictions. Thus, an approach with traces, or "silent copies" for that matter, is superior to an MC/MD-approach.

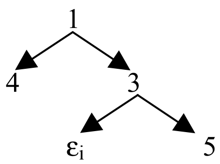
Of course, this does not invalidate any syntax-internal arguments in favor of an MC/MD-approach. However, it calls for extra assumptions on the PF-branch of C_{HL}. An *ad hoc* solution would be to exceptionally violate the principle of inclusiveness and throw in empty pointers at the stage where linear precedence is computed. Take an MC/MD-graph like (195).

(195)



The idea is to transform (195) into the constituent structure tree-like graph in (196).

(196)



The "empty pointer" in (196) can then play the role of a trace in the solution of Jacobson's problem. Roughly, the transition from (195) to (196) requires a substitution in the "residual" ID and SP, RID and RSP. These are respectively defined as ID - ID^{φ} and SP - SP^{φ}. Thus, assume for (195) that SP = { <4,3>, <4,5> } and that <1,4> is a strong link. We thus derive the following sets.

(197) a. ID^{φ} = {<1,4>, <1,3>, <3,5>}
b. RID = {<3,4>}
c. SP^{φ} = {<4,3>}
d. RSP = {<4,5>}

Substitution would replace 4 in RID and RSP by an empty pointer, yielding ID^{ε} = {<3,ε_i>} and SP^{ε} = {<ε_i,5>}.[125] Projection of P will then, instead of defining φ-Precedence as developed in section 3.4.3, have to be modified as follows. ($\Downarrow^{\cup}$ is interpreted in terms of $ID^{\varphi} \cup ID^{\varepsilon}$).

(198) ε-Precedence

a. $(\forall x,y \in N)[\ \langle x,y\rangle \in SP^{\varphi} \cup SP^{\varepsilon} \rightarrow \langle x,y\rangle \in P\ ]$

b. $(\forall x,y,z,w \in N)[\ (\ \langle x,y\rangle \in P \wedge z \in \Downarrow^{\cup}x \wedge w \in \Downarrow^{\cup}y\ ) \rightarrow \langle z,w\rangle \in P\ ]$

The result for (195)/(196) is given in (199).

(199) $P = \{\langle 4,3\rangle, \langle \varepsilon_i,5\rangle, \langle 4,\varepsilon_i\rangle; \langle 4,5\rangle\}$

(199) encodes (196) at the level of P, which is what we set out to get.

Let me finally address the following fairly technical question: To what extent does the hybrid operation DoIC/DoID obliterate fundamental differences between the operations Merge and Move replaced by it? Or put differently: Is there a difference between base positions and derived positions which cannot be adequately captured in MC/MD-syntax? Clearly, this question is closely related to the reconstruction of Θ-theory within minimalist syntax.[126]

[125] The technical details of substitution can be devoloped roughly along the lines of "(pre-) terminalization" via Select, given in section 3.3.3. Thus, we have to allow an additional item into the realm of pointers, namely, the empty pointer <∅,[∅]>. We must then set up a "substitution structure" SS = <ε,RID,RSP>, where ε is an empty pointer. The operation *Substitute*(x,y) would update an SS as follows.

(i) *Substitute*(x,y) applies to a pair of pairs of pointers and updates an SS = <ε,RID,RSP> to SS' = <ε',RID',RSP'>, such that

a. ε = <A,I>, A a set and I a multiset, and

b. x (= <α,β>) ∈ RID ∩ ID, and

c. y (= <γ,δ>) ∈ RSP ∩ SP, and

d. α ∈ IA(γ) ∩ IA(δ), and

e. β = γ or β = δ , and

f. ε' = <A¤A,I>, and

g. RID' = RID[<α,ε>/<α,β>], and

h. RSP' = { RSP[<ε,δ>/<γ,δ>], if β = γ ; RSP[<γ,ε>/<γ,δ>], if β = δ }

Each step creates a distinct empty pointer. Conditions (ib) and (ic) guarantee that substitution applies only to pairs that have not yet undergone substitution. Condition (id) ensures that for each link the corresponding sibling precedence is recorded on the substituted empty pointer as well. Note that this rather cumbersome procedure is the maximally general version of substitution. It remains to be seen how much of this will actually be needed for linguistic purposes. Thus, considerable simplification may be possible if only a subset of "discarded links" has to be preserved in terms of empty pointers.

[126] Cf. Chomsky (1995a, section 4.6) and Collins (1997, p.69ff).

The GB approach to Θ-theory makes a clear distinction in this domain.[127] There are base positions of arguments, projected by phrase-structure rules at D-structure. These must be Θ-positions. On the other hand, there are derived positions targeted via movement operations at S-structure. These cannot be Θ-positions. Such an approach implies a ban on movement into Θ-positions. This consequence of Θ-theory has been translated into minimalist syntax by assuming two things. First, creating a Θ-position does *not* lead to feature checking, and, secondly, movement is blocked *unless* it leads to feature checking.[128] This part of the theory can straightforwardly be carried over into MC/MD-syntax by a principle like (200).

(200) $\mathrm{DoID}(\alpha,\beta)$ is blocked if
a. it does not involve feature checking, and
b. $\exists\gamma[\ \langle\gamma,\beta\rangle \in \mathrm{ID}\ ]$

The more complicated part of the discussion concerns arguments inserted in non-Θ-positions. These may check off formal features properly but end up without a Θ-role. Unless extra assumptions are made, such cases would have to count as well-formed, even if "semantically deviant." There are two cases to consider, namely, insertion into strong positions and insertion into weak positions. Clearly, only the latter produces structures that every theory of syntax would have to rule out. Thus, (201) has to be prevented from being generated before Spell-Out.

(201) [$_{AgrSP}$ Kermit$_i$ [$_{AgrOP}$ beans [$_{VP}$ t_i likes]]]

A general way of blocking this kind of insertion would be to disallow feature-checking via insertion.[129] In the DoID-system, this would be stated as follows.

(202) $\mathrm{DoID}(\alpha,\beta)$ cannot involve feature-checking if
$\neg\exists\gamma[\ \langle\gamma,\beta\rangle \in \mathrm{ID}\ ]$

Unless (202) is relativized appropriately, such an approach requires additional assumptions about expletives like *there*, which are standardly taken to be involved in feature-checking when inserted.[130] One option would be to transfer expletive insertion into the PF-component.

A radical alternative to the approach just outlined would be to dispense with principles (200) and (202) and follow Stabler (1996, 1998) in allowing the checking of selectional features instead. The consequences of (200) and (202) could be preserved by further structuring FFORM, i.e. the formal features of each lexical item. In particular,

---

[127] See section 2.4 and references cited there.

[128] For an alternative to the first assumption, see Hornstein (1998) and Bošković (1994).

[129] Interestingly, in MC/MD-syntax the DP *beans* could not be linearly ordered wrt the remainder of the clause, given that its position is weak, i.e. $\langle AgrOP,DP\rangle \in WID$ and thus $\langle AgrOP,DP\rangle \notin ID^{\varphi}$. Cf. sections 3.3.1 and 3.4.3.

[130] Cf. Chomsky (1995a, p.286f).

the checking of selectional (S-)features has to be a precondition on the checking of the familiar “offensive“ O-features. In addition, principle (203), which also constitutes part of one approach to X°-movement entertained in section 3.4.2, has to be assumed.

(203) Checking Condition on DoID
Every application of DoID involves feature checking

Inserting an argument α into a non-Θ-position would now be blocked, because the S-features of α have not been checked. This in turn prevents the O-features of α from being accessible for checking. Consequently, (203) is violated. Given this, movement of an argument into Θ-position is likewise ruled out because arguments that have already been inserted will have their S-features checked and are thus unable to check against another selector. Once again (203) is violated. Expletives, on this view, could be treated as an exception, insofar as they do not possess S-features. They can thus be inserted in non-Θ-position right away.[131]

To the extent that the latter approach is desirable, it provides the proper conditions for my (conditional) hypothesis 2.[132]

H2: Under certain conditions, (part of) the theta criterion has to be implemented as a syntactic well-formedness constraint

---

[131] Under this perspective, adjuncts could be taken to possess a categorially more or less specified S-feature, checked off against the appropriate constituent on insertion. This would have to be asymmetric checking, leaving the features of the adjunction host unaltered. Taking such “adjunct-features“ together with “adjunction-features,“ which arguably drive operations like scrambling (cf. Grewendorf&Sabel 1997), to constitute the set of “A-features,“ a general approach to adjunction in MC/MD-syntax could look as follows. In analogy to WID for covert movement, an “adjoined immediate dominance“ relation, AID, can be postulated. Derivational stages have to be enriched accordingly to $\langle\mathbb{R}$, ID, WID, AID$\rangle$-quadruples. The update of AID, brought about by DoID(α,β), is given in (i).

(i) $AID' = \begin{cases} AID \cup \{\langle\langle A¤A,I\rangle,\beta\rangle\}, \text{ if } DoID(\alpha,\beta) \text{ involves checking of A-features} \\ AID, \text{ otherwise} \end{cases}$

In terms of AID, it is for example possible to formulate an adjunct island constraint as follows.

(ii) Adjunct Island Condition
DoID(α,β) is blocked if
$\exists\gamma,\delta[\ \langle\gamma,\delta\rangle \in AID \wedge \langle\delta,\beta\rangle \in ID^{+}\ ]$

Further notions of adjunction structures should likewise be definable. See Kolb (1997b) and Kracht (1999). One advantage of the AID-approach is that it obviates the need for using specialized labels and the concomitant specialized Merge operation, discussed in section 2.6.3.

[132] Adopting H2, even unconditionally, does not imply any strong commitment to how Θ-theory is conceived of semantically. Thus, each lexical item could be associated with an arbitrary selection grid by the specification operation that maps the lexicon into LEX. The resulting syntactic structures could then be evaluated against the semantics of the lexical item whenever such evaluation seems appropriate and by whatever theoretical mechanism deemed adequate.

Note also that a counterpart of (203) is Stabler's (1996, 1998) way of radicalizing "resource sensitivity" of his system. Thus, in standard minimalist syntax only Move is taken to observe a "last resort" condition, such that it only applies if it leads to feature-checking, i.e. if it is "triggered." Although more restrictive than GB, where Move-α applies freely, this leaves open the question which force Merge is driven by. The answer tends to appeal to something like "virtual conceptual necessity," according to which no structures could be built without Merge.[133] In MC/MD-syntax, the "Merge-part" of DoIC/DoID is driven by the Single Root Condition, repeated in (24) (cf. 3.3).

(204) SYN = Σ (= $\langle\langle \mathbb{N}_T, \mathbb{N}_P\rangle$,ID,WID$\rangle$) is translatable into PHON/SEM, only if
$(\exists\alpha \in dom(\text{ID}))(\forall\beta \in \mathbb{N}_P \cup dom(\text{ID}) \cup ran(\text{ID}))[\ \langle\alpha,\beta\rangle \in \text{ID}^*\ ]$

To the extent that further unification of Merge and Move, and thus further "dehybridization" of DoIC/DoID is desirable, Stabler's radical resource sensitivity, as expressible in (203), looks like a promising way to go. Again, under this condition, H2 is vindicated.

3.6 Summary

Section 3 has presented "MC/MD-syntax," i.e. the syntax of multiconstituency and multidominance, in response to the specific problem situation and results arrived at in section 2 (cf. 2.9). The first preparatory step (cf. section 3.1) has consisted in limiting the scope of the theory to "narrow syntax." This step crucially reduces the interface burden involved in establishing syntactic well-formedness. The minimalist Principle of Full Interpretation (FI) is recast as a filter on "uninterpretable" features, called "offensive" (O-)features here. I repeat the definitions of well-formedness ("grammaticalness") and the "Principle of Translatability," which replaces FI, in (205) and (206).

(205) A linguistic expression L (= <phon,syn,sem>) is *grammatical* iff
a. there is a derivation D that generates *syn*, and
b. *syn* can be translated into *phon* and *sem*, and
c. *phon* and *sem* are well-formed

(206) Principle of Translatability
Syn is not translatable into *phon*/*sem* if *syn* contains unchecked O-features

This rather weak version of FI goes along with the absence of any appeal to interpretation-induced "invisibility" of syntactic elements, the latter in fulfillment of hypothesis 4.

---

[133] Cf. Collins (1997) for detailed discussion of whether this is a result of economy principles.

Section 3.2 introduces the graph-theoretic background for MC/MD-syntax. Thus, while standard constituent structure trees imply the Single Mother Condition (SMC), repeated in (207), this condition does not hold for graphs allowing MC/MD.

(207) Single Mother Condition (SMC)
$(\forall x,y,z \in N)[\ (\ xIDz \wedge yIDz\ ) \rightarrow x = y\ ]$

A very general variant of graphs for implementing MC/MD is introduced under the name of "rooted MC/MD-graphs."

(208) A *rooted MC/MD-graph* is a pair $\langle N,ID \rangle$, where
N is a finite set, the set of nodes,
ID is a binary relation on N, the immediate dominance relation,
and such that
(1) ID^+ is irreflexive (Acyclicity)
(2) $(\exists x \in N)(\forall y \in N)[\ \langle x,y \rangle \in ID^*\ ]$ (Single Root Condition)

Rooted MC/MD-graphs serve as background objects for the derivational system of MC/MD-syntax. For linguistic purposes, two main restrictions have to be imposed on these graphs. First, given that MC/MD is intended to replace copies and chains in the implementation of movement operations, the C-Command Condition on movement must be guaranteed. This is the role of the "Connected Ancestor Condition" (CAC).

(209) Connected Ancestor Condition (CAC)
$(\forall x,y,z \in N)[\ (\ \langle x,z \rangle \in ID \wedge \langle y,z \rangle \in ID\ ) \rightarrow (\ \langle x,y \rangle \in ID^* \vee \langle y,x \rangle \in ID^*\ )\ ]$

Secondly, it has to be shown how linear precedence can be handled in MC/MD-structures. Section 3.2 discusses two (partial) techniques for achieving this. The first one projects linear precedence from obligatorily ordered terminals into "orderable" regions of the rooted MC/MD-graph. The crucial definitions are repeated in (210) and (211). They are constraints on "ordered rooted MC/MD-graphs."

(210) Obligatory Ordering of Terminals
$(\forall x,y \in T)[\ x \neq y \rightarrow (\ \langle x,y \rangle \in P \vee \langle y,x \rangle \in P\ )\ ]$

(211) (Precedence Inheritance Condition)
$(\forall x,y \in N)[\ \langle x,y \rangle \in P \leftrightarrow \langle \Downarrow x,\Downarrow y \rangle \in \mathbb{P}\ ]$

Roughly, non-terminals can be ordered wrt each other in terms of P if they dominate continuous constituents.

The second technique linearly orders nodes that have a common parent, i.e. "siblings." This ordering can be projected into the graph after the surface position for each constituent has been determined and redundant edges, or "links," have been removed. This is the approach advocated for "phrase-linking grammar" (PLG)

(Peters&Ritchie 1981), which section 3.2 goes on to discuss. The “linked trees“ of PLG differ from rooted MC/MD-graphs in two main respects. First, the ID-relation is partitioned into *I* and *L*, where *I* determines the surface positions and *L* records links for “abstract positions.“ Secondly, a counterpart of the CAC is built into the definition of linked trees.

Section 3.2 closes off with an excursus on “structure-sharing,“ (3.2.1), as familiar from frameworks like HPSG (Pollard&Sag 1987, 1994). Among other things this excursus illustrates the importance of indexation in dealing with *type*- vs. *token*-identity in feature structures.

The central section 3.3 develops the core of MC/MD-syntax. This comes in two closely related variants, the “DoIC-system“ (3.3.1) and the “DoID-system“ (3.3.3). These are respectively named after their main syntactic operations, “DoIC“ and “DoID,“ which themselves refer to “immediate constituency“ (IC) and “immediate dominance“ (ID), respectively. Both systems coincide in that their output essentially corresponds to a subset of rooted MC/MD-graphs, which is brought about by the binary operation DoIC/DoID, defining an IC-/ID-relation on syntactic elements. This hybrid operation does double duty, replacing both Merge and Move of minimalist syntax. MC/MD results from the possibility of individual syntactic elements being an argument of that operation more than once. In fulfillment of hypothesis 3, both systems imply the “Strictest Cycle Condition,“ given in (212).

(212) Strictest Cycle Condition
Every syntactic operation creates a root node

This is mainly guaranteed by a counterpart to the CAC above, called “Ancestor Condition on DoIC/DoID.“ (213) and (214) provide the DoIC-variant of the system.

(213) *DoIC*(α,β) is a syntactic operation updating a derivational stage Σ (= $<\mathbb{N}, IC, WIC>$) to a derivational stage Σ' (= $<\mathbb{N}', IC', WIC'>$), such that
a. α and β are terms from $\mathbb{N} \cup dom(IC) \cup ran(IC)$, $\alpha \neq \beta$, and
b. $\mathbb{N}' = \mathbb{N} - \{\alpha,\beta\}$, and
c. $IC' = IC \cup \{ <\alpha,<\alpha,\beta>>, <\beta,<\alpha,\beta>> \}$, and
d. $WIC' = \begin{cases} WIC \cup \{ <\beta,<\alpha,\beta>> \}, \text{ if DoIC}(\alpha,\beta) \text{ involves checking of a weak feature} \\ WIC, \text{ otherwise} \end{cases}$

(214) Ancestor Condition on DoIC
$\forall\alpha,\beta$, DoIC(α,β) can only apply if
a. $\neg\exists\gamma\, [<\alpha,\gamma> \in IC]$, and
b. $\forall\delta\, [<\beta,\delta> \in IC \rightarrow <\delta,\alpha> \in IC^*]$

I assume the notational convention that for DoIC/DoID(α,β), α is the “projecting“ element. (214a) ensures that that element is undominated when DoIC applies. Also, given (214b), the ancestors of the non-projecting element β will be “connected“ in terms of IC*. This means that multiply-linked constituents are immediately dominated by exact counterparts of those nodes the multiple copies in a chain observing C-Command would be immediately dominated by. Lowering and sideward movement is prohibited.

The resource set of lexical items $\mathbb{N}$ (="numeration") is carried over from minimalist syntax and (213b) ensures that it is successively emptied.

Given strictest cyclicity, covert movement is dealt with on the same cycle as overt movement. This is implemented in terms of the WIC/WID-relation, (W = "weak"), keeping a record of links created by the checking of weak features. In order to prevent constituents from surfacing in covert positions, the mechanism generating *phon* must discard "weak links." Likewise, among multiple strong links all but the most prominent one must be eliminated as well. The structure arising from this reduction procedure is called IC^{φ} (ID^{φ}), defined in (215).

(215) $\forall\alpha,\beta, <\alpha,\beta> \in IC^{\varphi}$ iff
a. $<\alpha,\beta> \in$ IC-WIC, and
b. $\neg\exists\gamma[\, <\alpha,\gamma> \in \text{IC-WIC} \wedge <\beta,\gamma> \in (\text{IC-WIC})^{+}\,]$

The DoID-system (3.3.3) then extends the DoIC-system in response to potential problems inherent in the latter. These are discussed in section 3.3.2. In particular, a hypothetical version of minimalist syntax is considered that interprets multiple occurrences of constituents in structures like (216) not as copies of a chain but as a single object.

(216) $\{ \alpha, \{ \delta, \{ \gamma, \{ \alpha, \beta \} \} \} \}$

An appeal to set-theoretic extensionality would allow that. Such an approach could be seen as a direct method of deriving MC/MD-structures. However, it seems that such a move hides away the complexity of substitution in multiple occurrences and keeping track of identical elements in a structure in the meta-language. Yet, given the indiscriminate use of terms in the DoIC-system as well, this problem would carry over into MC/MD-syntax.

As a remedy, the DoID-system assigns checking resources to a single location, confining them to "terminal nodes." This is the domain defined over "lexical items." On top of this, a domain of O-feature-free "pointers" is postulated. These pointers function as non-terminals nodes in MC/MD-graphs, providing the input to DoID in the way described above. At the same time, indexation, similar to labeling in minimalist syntax, allows keeping track of which terminals correspond to which pointers, so that checking resources can be eliminated in the right place. Thus, apart from an indexation procedure compatible with the minimalist principle of inclusiveness, the DoID-system adds a second numeration for terminals, $\mathbb{N}_T$, to derivational stages. $\mathbb{N}_T$ is "updated" by "substituting" the terminal heads of pointers α and β if DoID(α,β) involves feature checking. This is given in (217). (*A[x/y]* denotes substitution of *x* for *y* in *A*.)

(217) $$\mathbb{N}_T' = \begin{cases} \mathbb{N}_T[th(\alpha)'/th(\alpha),\ th(\beta)'/th(\beta)], & \text{if DoID}(\alpha,\beta) \text{ involves feature checking} \\ \mathbb{N}_T, & \text{otherwise} \end{cases}$$

Section 3.3 includes a brief comparison of MC/MD-syntax with closely related approaches to cyclicity such as Bobaljik (1995a), Groat&O'Neil (1996), and Stabler (1996, 1998).

Section 3.4 develops MC/MD-syntax further in the areas of C-Command, X°-movement, and linear precedence. Section 3.4.1 shows that the version of C-Command most successful in defining a notion like “closeness“ on complex structures involving remnant movement is “Weak C-Command,“ given in (218).

(218) Weak C-Command (WCC)
α *weakly c-commands* β, iff
a. $\alpha \neq \beta$, and
b. $<\alpha,\beta> \notin$ SD, and
c. $(\exists\gamma \in \text{IA}(\alpha))(\forall\delta \in \text{IA}(\beta))[<\gamma,\delta> \in \text{ID}^*]$

Crucially, C-Command in MC/MD-graphs has to take into account the set of immediate ancestors of a given node *x*, “IA(x).“ Likewise, since “extraction“ in MC/MD-syntax does not mean removal of a (set of) node(s) from a constituent, but “linking“ it to the outside, the dominance condition involved in C-Command for trees has to be replaced by a “strict dominance“ (SD) condition in WCC, i.e. (218b). The definition of SD is given in (219).

(219) Strict Dominance (SD)
$\alpha\text{SD}\beta \Leftrightarrow_{\text{def}} \forall\gamma[\ \gamma \in \text{IA}(\beta) \rightarrow <\alpha,\gamma> \in \text{ID}^*]$

Section 3.4.2 discusses four ways of dealing with X°-movement in MC/MD-syntax compatible with strictest cyclicity. These are sideward movement, as proposed by Bobaljik (1995a) and Bobaljik&Brown (1997), “Münchhausen-style“ movement, i.e. movement that projects its own target (cf. Ackema et al. 1993), X°-movement into (multiple) specifiers (cf. Toyoshima 1997, Fukui&Takano 1998), and “no movement,“ the latter relegating spell-out of heads to lexical specification and syntactic checking of selectional features. The first two approaches require a relativization of the Ancestor Condition. The last three approaches call for richer lexico-morphological assumptions. I leave it as an open question which approach should ultimately be pursued.

Section 3.4.3 shows how “sibling precedence“ (SP) can be made a part of the operation DoIC/DoID. This involves the addition of an SP-relation to derivational stages and must be supplemented with projection techniques from section 3.2.

Finally, section 3.5 addresses three kinds of objection to MC/MD-syntax, sketching ways how to meet them. Firstly, to the extent that overt movement leads to a “freezing-effect“ for covert extraction, it may not be obvious how this is dealt with on a single cycle. However, given the availability of the WIC/WID-relation, such effects can be captured as in (220).

(220) Ban on Pseudo-Remnant Movement
DoID(α,β) is blocked if
$\exists\gamma,\delta[\ <\gamma,\delta> \in \text{WID} \wedge <\gamma,\beta> \in \text{ID}^+ \wedge <\beta,\delta> \in \text{ID}^+\ ]$

Secondly, there seem to be cases where a single constituent has effects at various positions in a *phon*-representation. The first one concerns German WH-copy constructions. Here, I speculate that, given severe morphological constraints on the copiable items, a lexico-morphological solution involving complementizers may be found. Other cases indicate that displaced overt constituents are accompanied by something like an empty terminal or "trace" in their base position. The familiar phenomenon of *wanna*-contraction, however, does not necessitate any extra-assumptions since its main effects must be stated wrt hierarchical rather than linear notions. Yet, a subtle fact from VP-deletion pointed out by Jacobson (1982) may require insertion of empty terminals after all. A substitution method is presented that transforms MC/MD-graphs into trees containing empty terminals at locations of "discarded links." It remains to be seen whether the full power of such a "worst case" remedy is really required.

Finally, the more general question is raised whether a hybrid DoIC/DoID-approach obliterates irreducible differences between Merge and Move, ultimately stemming from irreducible properties of the base vs. the transformational component of generative syntax. This most urgently calls for a method to prevent movement into Θ-positions and argument insertion into non-Θ-positions. In turn a decision on whether to allow selection-phenomena into the realm of feature-checking must be made. If that is not allowed, the above phenomena can be directly blocked by constraints on the application of DoIC/DoID. If, on the other hand, one allows, and requires, the checking of selectional (S-)features, another picture emerges. According to such a view, one may stipulate (221) as a general constraint on DoIC/DoID.

(221) Checking Condition on DoID
Every application of DoID involves feature checking

In addition one has to make S-feature elimination a precondition on O-feature checking by giving lexical items a more elaborate internal structure. This approach to recapturing asymmetries between Merge and Move would constitute a condition under which hypothesis H2 comes into play, as would the "no movement" approach to X°-movement (3.4.2).

4 Conclusion

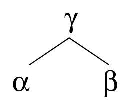
Ist denn der Weg so lang?
(Georg Büchner, Leonce und Lena)

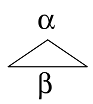
In this study, I have most importantly argued that the minimalist operations “Merge“ and “Move,“ as chiefly presented in Chomsky (1995a), should be unified into a single, hybrid operation called “DoID“ (or, alternatively, “DoIC“)(cf. section 3.3). This responds to the fact that both the binary (“generalized“) transformation Merge and the singulary transformation Move add structure at the root of the objects transformed. This is given schematically in (1) and its corresponing graph-theoretic form in (2) and (3).

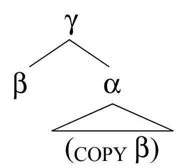
(1) a. $BT(\alpha,\beta)$
$\alpha \quad \beta \rightarrow (\alpha, \beta)$

b. $ST(({}_{\alpha}\ \beta\ ))$
$({}_{\alpha}\ \beta) \rightarrow (\beta, ({}_{\alpha}\ (\text{COPY}\ \beta)))$

(2) a. $\alpha \qquad \beta$

b. 

(3) a. 

b. 

The reconstruction of these cases in terms of DoID brings out an essentially identical behavior, in that both operations “update“ an immediate dominance (ID-)relation in exactly the same way. This is stated in (4).

(4) $DoID(\alpha,\beta);\ ID' = ID \cup \{ <\gamma,\alpha>, <\gamma,\beta> \}$

The “growth effect“ of syntactic operations has in earlier terminology been called “cyclicity.“ In this study, I postulate that cyclicity should be absolute. This is the content of hypothesis 3 (cf. 2.5.2, 2.7.2).

H3: There are no counter-cyclic syntactic operations

For the DoID-system, cyclicity in the sense of (5) is guaranteed by an "Ancestor Condition," given in (6). (For expository purposes, arguments of DoID(α,β) are ordered, such that α "projects.")

(5) Strictest Cycle Condition
Every syntactic operation creates a root node

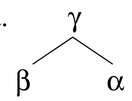
(6) Ancestor Condition on DoID
$\forall\alpha,\beta$, DoID(α,β) can only apply if
a. $\neg\exists\gamma[\langle\gamma,\alpha\rangle \in \text{ID}]$, and
b. $\forall\delta[\langle\delta,\beta\rangle \in \text{ID} \rightarrow \langle\alpha,\delta\rangle \in \text{ID}^*]$

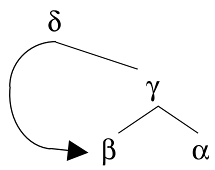
It is argued in section 2.5.2 (cf. 2.6.3, 2.7.2), that cyclicity is empirically desirable. In 3.3. and 3.4.2 it is shown in addition how formal obstacles to strictest cyclicity involving "covert movement" and "X°-movement" can be overcome.

The crucial difference between Merge and Move, as illustrated in (1)-(3), lies in the fact that the item that "moves," i.e. β, is already dominated when Move applies. In the DoID-system this means that there is a δ, such that $\langle\delta,\beta\rangle \in$ ID when DoID applies. Take the following two-step derivation.

(7) a. DoID(α,β); ID = $\{ \langle\gamma,\alpha\rangle, \langle\gamma,\beta\rangle \}$
b. DoID(γ,β); ID' = $\{ \langle\gamma,\alpha\rangle, \langle\gamma,\beta\rangle, \langle\delta,\gamma\rangle, \langle\delta,\beta\rangle \}$

Now consider the graphs corresponding to the respective ID-relations.

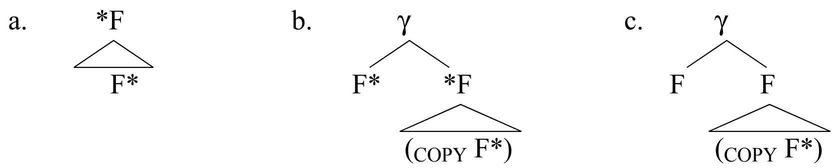
(8) a.


b.


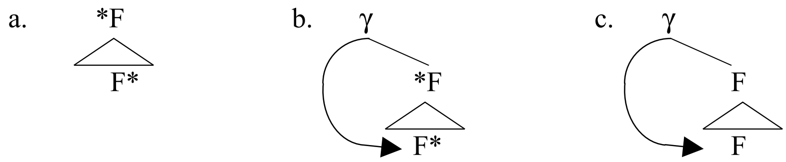
Applying DoID to the same node, β, twice, results in β being immediately dominated by more than one node. This property is called "multiconstituency" or "multidominance." It is the hallmark of "MC/MD-syntax" developed in section 3 of this study, of which the DoID-system is one instantiation. MC/MD-syntax responds to hypothesis 5, developed in section 2.

H5: The proper treatment of (unbounded) dependencies in minimalist syntax requires appeal to "multiconstituency"/"multidominance" (MC/MD)

One of the main arguments for MC/MD-syntax is its capability of restraining the "power of copying" (cf. section 2.7), the latter involved in Move, illustrated in (1b) and (3). In minimalist syntax, Move is taken to eliminate "checking resources" in local configurations. Uneliminated resources cause ill-formedness. Thus, a "functor," *F, “attracts“ an “argument“ F* into its domain, [F* *F], where they cancel, [F F]. The copying approach does not directly yield such a result, as the transition in (9) shows.

(9)


Copies retain checking resources. To remedy this kind of “resource paradox,“ appeal to “higher-order“ objects, called “chains,“ has been made in standard minimalist syntax (cf. 2.7). MC/MD-syntax, instead, disallows copying, as shown in (10).

(9)


The bulk of this study consists in analyzing the technical detail surrounding both the minimalist system (cf. section 2) and MC/MD-syntax (cf. section 3). Especially, the role of set-theoretic notions used for replacing/(re-)constructing graph-theory is discussed. A formal graph-theoretic background for MC/MD-syntax is provided in section 3.2.

This focusing in on detail, I argue, is necessary, if the overall minimalist concern, as expressed in Q1 (cf. section 1), is to be seriously addressed.

Q1: How “perfect“ is language?

Along the way, four additional hypotheses are defended. First, one way of treating X°-movement in MC/MD-syntax (cf. 3.4.2) together with the view that Merge and Move should be further approximated wrt “resource sensitivity“ (cf. 3.5) provides conditions under which (part of) the theta criterion from GB theory has to be implemented. This is formulated in the (conditional) hypothesis 2.

H2: Under certain conditions, (part of) the theta criterion has to be implemented as a syntactic well-formedness constraint

Such implementation involves addition to MC/MD-syntax of condition (11) as well as the association of lexical items with (arbitrary) selectional grids.

(11) Checking Condition on DoID
Every application of DoID involves feature checking

Secondly, the interface-oriented set-up of minimalist syntax is taken to involve a close relationship between interpretability of objects at the interface and accessiblity, or “visibility“ of such objects to syntax-internal operations. I have offered various arguments against this conception (cf. 2.6.4, 2.7.1), and consequently defended hypothesis 4.

H4: Invisibility of syntactic elements for C_{HL} has either to be avoided or it must be stipulated and properly implemented

Thirdly, I have argued that economy principles should play no role in minimalist syntax, as expressed in hypothesis 6.

H6: There are no economy principles operating in minimalist syntax

This has been met, first of all, by developing a system, i.e. MC/MD-syntax, which does without such principles. Secondly, I have offered an empirical argument against the most widely assumed version of an economy principle in standard minimalism (cf. 2.8).

Finally, in an attempt to develop an intuition about minimalist theory construction and its relation to structural notions, I have introduced the somewhat enigmatic hypothesis 1.

H1: The whole is more than the sum of its parts

In retrospect, I hope, H1 also applies to this study.

References

ABEILLÉ, Anne (1993): *Les Nouvelles Syntaxes*. Paris: Armand Colin.
ABRAHAM, Werner, Samuel EPSTEIN, Höskuldur THRÁINSSON & Jan-Wouter ZWART (eds.) (1996): *Minimal Ideas*. Amsterdam: John Benjamins.
ACKEMA, Peter, Ad NEELEMAN & Fred WEERMAN (1993): "Deriving Functional Projections." *NELS 23*: 17-31.
ACZEL, Peter (1988): *Non-Well-Founded Sets*. Stanford: CSLI-Publications.
ALEXIADOU, Artemis, Elena ANAGNOSTOPOULOU, Sjef BARBIERS & Hans-Martin GÄRTNER (forthcoming): "Introduction." in ALEXIADOU, Artemis et al. (eds.).
ALEXIADOU, Artemis, Elena ANAGNOSTOPOULOU, Sjef BARBIERS & Hans-Martin GÄRTNER (eds.)(forthcoming): *Dimensions of Movement*. Amsterdam: John Benjamins.
BAKER, Mark (1996): *The Polysynthesis Parameter*. Oxford: OUP.
BALTIN, Mark & Anthony KROCH (eds.) (1989): *Alternative Conceptions of Phrase Structure*. Chicago: University of Chicago Press.
BARBOSA, Pilar, Danny FOX, Paul HAGSTROM, Martha McGINNIS & David PESETSKY (eds.) (1998): *Is the Best Good Enough?* Cambridge, Mass.: MIT Press.
BARKER, Chris & Geoffrey K. PULLUM (1990): "A Theory of Command Relations." *Linguistics and Philosophy 13*: 1-34.
BARSS, Andrew (1995): "Extraction and Contraction." *Linguistic Inquiry 26.4*: 681-694.
BAYER, Josef (1984): "Comp in Bavarian Syntax." *The Linguistic Review 3*: 209-274.
BAYER, Josef (1996): *Directionality and Logical Form*. Dordrecht: Kluwer.
BERWICK, Robert C. (1994): "Cartesian Computation and Linguistics in a Current Context." in MARTÍN-VIDE, Carlos (ed.): *Current Issues in Mathematical Linguistics*. Amsterdam: Elsevier. 341-352.
BERWICK, Robert C. & Partha NIYOGI (1996): "Learning from Triggers." *Linguistic Inquiry 27.4*: 605-622.
BESTEN, Hans den & Gert WEBELHUTH (1990): "Stranding." in GREWENDORF, Günther & Wolfgang STERNEFELD (eds.): 77-92.
BIENENSTOCK, Elie & Stuart GEMAN (1995): "Compositionality in Neural Systems." in ARBIB, Michael A. (ed.): *The Handbook of Brain Theory and Neural Networks*. Cambridge, Mass.: MIT Press. 223-226.
BIERWISCH, Manfred (1963): *Grammatik des Deutschen Verbs*. Berlin: Akademie Verlag.
BIERWISCH, Manfred (1982): "Formal and Lexical Semantics." *Linguistische Berichte 80*: 3-17.
BIERWISCH, Manfred (1997): "Lexical Information from a Minimalist Point of View." in WILDER, Chris et al. (eds.): 227-266.
BLACKBURN, Patrick & Wilfried MEYER-VIOL (1996): "Modal Logic and Model Theoretic Syntax." in RIJKE, Maarten de (ed.): *Advances in Intensional Logic*. Dordrecht: Kluwer. 27-58.

BLEVINS, James (1990): *Syntactic Complexity*. Ph.D. Dissertation. University of Massachusetts at Amherst.
BLOOMFIELD, Louis A. (1997): *How Things Work*. New York: Wiley and Sons.
BLUTNER, Reinhard (1993): "Dynamic Generalized Quantifiers and Existential Sentences in Natural Languages." *Journal of Semantics 10*: 33-64.
BOBALJIK, Jonathan (1995a): "In Terms of Merge." *MITWPL 27*: 65-103.
BOBALJIK, Jonathan (1995b): *Morphosyntax*. Ph.D. Dissertation. MIT.
BOBALJIK, Jonathan & Samuel BROWN (1997): "Interarboreal Operations." *Linguistic Inquiry 28*: 345-356.
BOLINGER, Dwight (1979): "Pronouns in Discourse." *Syntax and Semantics 12*. New York: Academic Press. 291-309.
BORER, Hagit (1984): *Parametric Syntax*. Dordrecht: Foris.
BOŠKOVIĆ, Željko (1994): "D-Structure, Theta-Criterion, and Movement into Theta Position." *Linguistic Analysis 24*: 247-268.
BOŠKOVIĆ, Željko (1995): "Case Properties of Clauses and the Greed Principle." *Studia Linguistica 49*: 32-53.
BOUCHARD, Denis (1984): *On the Content of Empty Categories*. Dordrecht: Foris.
BOUCHARD, Denis (1986): "Empty Categories and the Contraction Debate." *Linguistic Inquiry 17.1*: 95-104.
BRESNAN, Joan (1982): "Control and Complementation." in BRESNAN, Joan (ed.): 282-390.
BRESNAN, Joan (ed.) (1982): *The Mental Representation of Grammatical Relations*. Cambridge, Mass.: MIT Press.
BRODY, Michael (1984): "On Contextual Definitions and the Role of Chains." *Linguistic Inquiry 15.3*: 355-380.
BRODY, Michael (1995): *Lexico-Logical Form*. Cambridge, Mass.: MIT Press.
BRODY, Michael (1998): "Projection and Phrase Structure." *Linguistic Inquiry 29.3*: 367-398.
BÜRING, Daniel (1996): "Economy." *Sprachwissenschaft in Frankfurt. Arbeitspapier Nr. 17*.
CARPENTER, Bob (1992): *The Logic of Typed Feature Structures*. Cambridge: CUP.
CHAMETZKY, Robert A. (1996): *A Theory of Phrase Markers and the Extended Base*. Albany: State University of New York Press.
CHOMSKY, Noam (1957): *Syntactic Structures*. The Hague: Mouton.
CHOMSKY, Noam (1963): "Formal Properties of Grammars." in LUCE, Duncan R. et al. (eds.): *Handbook of Mathematical Psychology Vol. II*. New York: Wiley and Sons. 323-418.
CHOMSKY, Noam (1965): *Aspects of the Theory of Syntax*. Cambridge, Mass.: MIT Press.
CHOMSKY, Noam (1971): "Deep Structure, Surface Structure, and Semantic Interpretation." in STEINBERG, Danny & Leon JAKOBOVITS (eds.): *Semantics*. New York: CUP. 183-216.
CHOMSKY, Noam (1972): "Some Empirical Issues in the Theory of Grammar." in PETERS, Stanley (ed.): 63-130.
CHOMSKY, Noam (1975/1955[1]): *The Logical Structure of Linguistic Theory*. New York: Plenum.
CHOMSKY, Noam (1977): "On WH-Movement." in CULICOVER, Peter et al. (eds.): *Formal Syntax*. New York: Academic Press. 71-132.
CHOMSKY, Noam (1980a): *Rules and Representations*. Oxford: Blackwell.
CHOMSKY, Noam (1980b): "Rules and Representations." *The Behavioral and Brain Sciences 3*: 1-61.
CHOMSKY, Noam (1980c): "On Binding." *Linguistic Inquiry 11*: 1-46.
CHOMSKY, Noam (1981): *Lectures on Government and Binding*. Dordrecht: Foris.
CHOMSKY, Noam (1982): *Some Concepts and Consequences of the Theory of Government and Binding*. Cambridge, Mass.: MIT Press.

CHOMSKY, Noam (1986a): *Knowledge of Language*. New York: Praeger.
CHOMSKY, Noam (1986b): *Barriers*. Cambridge, Mass.: MIT Press.
CHOMSKY, Noam (1990): "On Formalization and Formal Linguistics." *Natural Language and Linguistic Theory 8*: 143-147.
CHOMSKY, Noam (1991): "Linguistics and Cognitive Science." in KASHER, Asa (ed.): *The Chomskyan Turn*. Oxford: Blackwell. 26-53.
CHOMSKY, Noam (1995a): *The Minimalist Program*. Cambridge, Mass.: MIT Press.
CHOMSKY, Noam (1995b): "Bare Phrase Structure." in WEBELHUTH, Gert (ed.): 383-439.
CHOMSKY, Noam (1995c): "Language and Nature." *Mind 104*: 1-61.
CHOMSKY, Noam (1998): "Some Observations on Economy in Generative Grammar." in BARBOSA, Pilar et al. (eds.): 115-127.
CHOMSKY, Noam (2000): "Minimalist Inquiries." in MARTIN, Roger et al. (eds.): *Step by Step.* Cambridge, Mass.: MIT Press. 89-155.
CHOMSKY, Noam (2001): "Derivation by Phase." in KENSTOWICZ, Michael (ed.): *Ken Hale. A Life in Language*. Cambridge, Mass.: MIT Press. 1-52.
CHOMSKY, Noam & George A. MILLER (1963): "Introduction to the Formal Analysis of Natural Languages." in LUCE, Duncan R. et al. (eds.): *Handbook of Mathematical Psychology Vol. II*. New York: Wiley and Sons. 269-321.
COLLINS, Chris (1997): *Local Economy*. Cambridge, Mass.: MIT Press.
COLLINS, Chris (2001): "Economy Conditions in Syntax." in BALTIN, Mark & Chris COLLINS (eds.): *The Handbook of Contemporary Syntactic Theory*. Oxford: Blackwell. 45-61.
CORMEN, Thomas, LEISERSON, Charles & Ronald RIVEST (1990): *Introduction to Algorithms*. Cambridge, Mass.: MIT Press.
CORNELL, Tom (1999): "Representational Minimalism." in KOLB, Hans Peter & Uwe MÖNNICH (eds.): 301-339.
CULICOVER, Peter W. (1993): "Evidence Against ECP Accounts of the That-T Effect." *Linguistic Inquiry 24*: 557-561.
CURRY, Haskell B. (1961): "Some Logical Aspects of Grammatical Structure." in JAKOBSON, Roman (ed.): *Structure of Language and its Mathematical Aspects, Proceedings of the Twelfth Symposium in Applied Mathematics*. Providence R.I.: American Mathematical Society. 56-68.
DOWTY, David (1982): "Grammatical Relations and Montague Grammar." in JACOBSON, Pauline & Geoffrey K. PULLUM (eds.): 79-130.
DOWTY, David (1989): "On the Semantic Content of the Notion 'Thematic Role'." in CHIERCHIA, Gennaro et al. (eds.): *Properties, Types, and Meaning, II*. Dordrecht: Kluwer. 69-129.
EIGEN, Manfred & Ruthild WINKLER (1985): *Das Spiel*. München: Piper.
ENGDAHL, Elisabeth (1986): *Constituent Questions*. Dordrecht: Reidel.
EPSTEIN, Samuel (1999): "Un-Principled Syntax: The Derivation of Syntactic Relations." in EPSTEIN, Samuel & Norbert HORNSTEIN (eds.): 317-345.
EPSTEIN, Samuel, Erich GROAT, Ruriko KAWASHIMA & Hisatsugu KITAHARA (1998): *A Derivational Approach to Syntactic Relations*. Oxford: OUP.
EPSTEIN, Samuel & Norbert HORNSTEIN (eds.) (1999): *Working Minimalism*. Cambridge, Mass.: MIT Press.
ESSLER, Wilhelm K. (1982): *Wissenschaftstheorie I*. Freiburg: Alber.
FANSELOW, Gisbert & Sascha FELIX (1993): *Sprachtheorie II*. Tübingen: Francke (UTB).
FAUCONNIER, Gilles (1994): *Mental Spaces*. New York: CUP.
FERGUSON, K. Scott (1996): "Shortest Move and Object Case Checking." in ABRAHAM, Werner et al. (eds.): 97-111.
FIENGO, Robert (1977): "On Trace Theory." *Linguistic Inquiry 8*: 35-62.

FILLMORE, Charles (1963): “The Position of Embedding Transformations in a Grammar.“ *Word 19*: 208-231.

FODOR, Janet D. & Stephen CRAIN (1990): “Phrase Structure Parameters.“ *Linguistics and Philosophy 13*: 619-659.

FODOR, Jerry A. (1975): *The Language of Thought*. Cambridge, Mass.: Harvard University Press.

FRAMPTON, John (1997): “Expletive Insertion.“ in WILDER et al. (eds.): 36-57.

FRAMPTON, John & Sam GUTMANN (1999): “Cyclic Computation, A Computationally Efficient Minimalist Syntax.“ *Syntax 2.1*: 1-27.

FRANK, Robert (1992): *Syntactic Locality and Tree Adjoining Grammar*. Ph.D. Dissertation. University of Pennsylvania.

FRANK, Robert & Shyam KAPUR (1996): “On the Use of Triggers in Parameter Setting.“ *Linguistic Inquiry 27.4*: 623-660.

FRANK, Robert & Anthony KROCH (1995): “Generalized Transformations and the Theory of Grammar.“ *Studia Linguistica 49*: 103-151.

FREGE, Gottlob (1987/1884[1]): *Die Grundlagen der Arithmetik*. Stuttgart: Reclam.

FREIDIN, Robert (1978): “Cyclicity and the Theory of Grammar.“ *Linguistic Inquiry 9.4*: 519-549.

FUKUI, Naoki & Yuji TAKANO (1998): “Symmetry in Syntax.“ *Journal of East Asian Linguistics 7*: 27-86.

GÄRTNER, Hans-Martin (1998): “Review of Jairo Nunes ‘The Copy Theory of Movement and Linearization of Chains in the Minimalist Program‘.“ *GLOT International 3.8*: 16-19.

GÄRTNER, Hans-Martin (1999): “Phrase-Linking Meets the Minimalist Program.“ BIRD, Sonya et al. (eds.): *WCCFL XVIII*: 159-169.

GAZDAR, Gerald, Ewan KLEIN, Geoffrey K. PULLUM & Ivan A. SAG (1985): *Generalized Phrase Structure Grammar*. Oxford: Blackwell.

GIBSON, Edward & Kenneth WEXLER (1994): “Triggers.“ *Linguistic Inquiry 25.3*: 407-454.

GOLDSCHLAGER, Les & Andrew LISTER (1988): *Computer Science*. New York: Prentice Hall.

GOODMAN, Nelson (1977/1951[1]): *The Structure of Appearance*. Dordrecht: Reidel.

GORDON, James Edward (1978): “Structures.“ London: Penguin.

GREFE, Wolfgang & Marcus KRACHT (1996): “Adjunction Structures and Syntactic Domains.“ *Arbeitspapiere des SFB 340*. Nr.80. Stuttgart&Tübingen.

GREWENDORF, Günther (1988): *Aspekte der Deutschen Syntax*. Tübingen: Narr.

GREWENDORF, Günther (1990): “Verbbewegung und Negation im Deutschen.“ *GAGL 30*: 57-125.

GREWENDORF, Günther (1995): *Sprache als Organ, Sprache als Lebensform*. Frankfurt/M.: Suhrkamp.

GREWENDORF, Günther & Joachim SABEL (1997): “Wh-Scrambling in the Minimalist Framework.“ in WILDER, Chris et al. (eds.): 58-80.

GREWENDORF, Günther & Wolfgang STERNEFELD (eds.) (1990): *Scrambling and Barriers*. Amsterdam: John Benjamins.

GRIMSHAW, Jane (1991): “Extended Projections.“ Ms. Rutgers University.

GROAT, Eric & Wayne O'NEIL (1996): “Spell-Out at the LF Interface.“ in ABRAHAM, Werner et al. (eds.): 113-139.

GUILFOYLE, Eithne & Máire NOONAN (1988): “Functional Categories and Language Acquisition“. Ms. McGill University.

HAEGEMAN, Liliane (1992): *Introduction to Government and Binding Theory*. Oxford: Blackwell.

HAEGEMAN, Liliane & Henk van RIEMSDIJK (1986): “Verb Projection Raising, Scope, and the Typology of Rules Affecting Verbs. *Linguistic Inquiry 17.3*: 417-466.

HAIDER, Hubert (1990): “ Topicalization and other Puzzles of German Syntax.“ in GREWENDORF, Günther & Wolfgang STERNEFELD (eds.): 93-112.

HAIDER, Hubert (1992): “Branching and Discharge.“ *Arbeitspapiere des SFB 340. Nr.23*. Stuttgart&Tübingen.
HAIDER, Hubert (1997): “Economy in Syntax is Projective Economy.“ in WILDER et al. (eds.): 205-226.
HAIDER, Hubert & Klaus NETTER (eds.)(1991): *Representation and Derivation in the Theory of Grammar*. Dordrecht: Kluwer.
HALL, Barbara (1964): “Review of Šaumjan & Soboleva (1963).“ *Language 40*: 397-410.
HALLE, Morris & Alec MARANTZ (1993): “Distributed Morphology and the Pieces of Inflection.“ in HALE, Kenneth & Samuel J. KEYSER (eds.): *The View from Building 20*. Cambridge, Mass.: MIT Press.
HARRIS, Zellig (1981): *Papers on Syntax*. Dordrecht: Reidel.
HAUGELAND, John (1981): “Semantic Engines.“ in HAUGELAND, John (ed.): 1-34.
HAUGELAND, John (ed.)(1981): *Mind Design*. Cambridge, Mass.: MIT Press.
HENY, Frank (1979): “A Review of Chomsky, ‘The Logical Structure of Linguistic Theory‘.“ *Synthese 40.2*: 317-352.
HEYCOCK, Caroline (1995): “Asymmetries in Reconstruction.“ *Linguistic Inquiry 26.4*: 547-570.
HIGGINBOTHAM, James (1985a): “On Semantics.“ *Linguistic Inquiry 16*: 547-593.
HIGGINBOTHAM, James (1985b): “A Note on Phrase Markers.“ *MITWPL Vol.6*: 87-101.
HODGES, Wilfrid (1983): “Elementary Predicate Logic.“ in GABBAY, Dov & Franz GUENTHNER (eds.): *Handbook of Philosophical Logic, Vol.1*. Dordrecht: Kluwer. 1-131.
HOEKSEMA, Jack (1985): “Traces in COMP.“ *CLS 21*: 160-170.
HORNSTEIN, Norbert (1998): “Movement and Chains.“ *Syntax 1.2*: 99-127.
HORNSTEIN, Norbert (2000): “On A-Chains: A Reply to Brody.“ *Syntax 3.2*: 129-143.
HORNSTEIN, Norbert & Amy WEINBERG (1995): “The Empty Category Principle.“ in WEBELHUTH, Gert (ed.): 241-296.
HORROCKS, Jeffrey (1987): *Generative Grammar*. Harlow: Longman.
HUANG, C.T. James (1995): “Logical Form.“ in WEBELHUTH, Gert (ed.): 125-175.
HUSSERL, Edmund (1980/1928[1]): *Vorlesungen zur Phänomenologie des Inneren Zeitbewußtseins*. Tübingen: Niemeyer.
IATRIDOU, Sabine (1990): “About Agr(P).“ *Linguistic Inquiry 21*: 551-577.
JACKENDOFF, Ray S. (1972): *Semantic Interpretation in Generative Grammar*. Cambridge, Mass.: MIT Press.
JACKENDOFF, Ray S. (1990): *Semantic Structures*. Cambridge, Mass.: MIT Press.
JACKENDOFF, Ray S. (1997): *The Architecture of the Language Faculty*. Cambridge, Mass.: MIT Press.
JACOB, Pierre (1986): “Les Représentations Logiques Sont-Elles Ambiguës?“ in RONAT, Mitsou & Daniel COUQUAUX (eds.): *La Grammaire Modulaire*. Paris: Minuit.
JACOBS, Joachim, Arnim von STECHOW, Wolfgang STERNEFELD & Theo VENNEMANN (eds.)(1993): *Syntax: Ein Internationales Handbuch*. Berlin: De Gruyter.
JACOBSON, Pauline (1982): “Evidence for Gaps.“ in JACOBSON, Pauline & Geoffrey K. PULLUM (eds.): 187-228.
JACOBSON, Pauline & Geoffrey K. PULLUM (eds.) (1982): *The Nature of Grammatical Representation*. Dordrecht: Reidel.
JAKOBSON, Roman (1974/1939[1]): “Das Nullzeichen.“ in *Aufsätze zur Linguistik und Poetik*. München: Ullstein. 44-53.
JOHNSON, David & Shalom LAPPIN (1997): “A Critique of the Minimalist Program.“ *Linguistics and Philosophy 20*: 273-333.

JOHNSON, David & Shalom LAPPIN (1999): *Local Constraints vs. Economy*. Stanford: CSLI-Publications.
JOHNSON, Mark (1988): *Attribute-Value Logic and the Theory of Grammar*. Stanford: CSLI-Publications.
JOSHI, Aravind (1985): "Tree Adjoining Grammars." in DOWTY, David et al. (eds.): *Natural Language Parsing*. New York: CUP. 206-250.
KAMBARTEL, Friedrich (1968): *Erfahrung und Struktur*. Frankfurt/M.: Suhrkamp.
KAPLAN, Ronald & Annie ZAENEN (1989): "Long-Distance Dependencies, Constituent Structure, and Functional Uncertainty." in BALTIN, Mark & Anthony KROCH (eds.): 17-42.
KARTTUNEN, Lauri & Martin KAY (1986): "Structure Sharing with Binary Trees." in SHIEBER, Stuart et al.: Part II chapter 1. 5-16.
KAYNE, Richard (1994): *The Antisymmetry of Syntax*. Cambridge, Mass.: MIT Press.
KEMPSON, Ruth (1988): "On the Grammar-Cognition Interface: The Principle of Full Interpretation." in KEMPSON, Ruth (ed.): *Mental Representations*. Cambridge: CUP. 199-225.
KENSTOWICZ, Michael (1994): *Phonology in Generative Grammar*. Cambridge, Mass.: Blackwell.
KISS, Katalin É. (ed.) (1995): *Discourse Configurational Languages*. Oxford: OUP.
KITAHARA, Hisatsugu (1995): "Target α: Deducing Strict Cyclicity from Derivational Economy." *Linguistic Inquiry 26*: 47-77.
KITAHARA, Hisatsugu (1997): *Elementary Operations and Optimal Derivations*. Cambridge, Mass.: MIT Press.
KITAHARA, Hisatsugu (2000): "Two (or More) Syntactic Categories vs. Multiple Occurrences of One." *Syntax 3.3*: 151-158.
KNERR, Richard (1991): *Mathematik*. Stuttgart: Parkland.
KOIZUMI, Masatoshi (1995): *Phrase Structure in Minimalist Syntax*. Ph.D. Dissertation. MIT
KOLB, Hans-Peter (ed.) (1997): "GB Blues. Two Essays on Procedures and Structures in Generative Syntax." *Arbeitspapiere des SFB 340. Nr.110*. Stuttgart&Tübingen.
KOLB, Hans-Peter (1997a): "Is I-Language a Generative Procedure?" in KOLB, Hans Peter (ed.): 3-11.
KOLB, Hans-Peter (1997b): "Categories, Segments, and Other Dangerous Things." in KOLB, Hans-Peter (ed.): 13-26.
KOLB, Hans-Peter & Uwe MÖNNICH (eds.) (1999): *The Mathematics of Syntactic Structure*. Berlin: Mouton de Gruyter.
KOOPMAN, Hilda & Anna SZABOLCSI (2000): *Verbal Complexes*. Cambridge, Mass.: MIT Press.
KORNAI, András & Geoffrey K. PULLUM (1990): "The X-Bar Theory of Phrase Structure." *Language 66*: 24-50.
KOSTER, Jan (1987): *Domains and Dynasties*. Dordrecht: Foris.
KRACHT, Marcus (1995): "Syntactic Codes and Grammar Refinement." *Journal of Logic, Language, and Information 4*: 41-60.
KRACHT, Marcus (1997): "Chapter 4." Ms. Free University Berlin.
KRACHT, Marcus (1999): "Adjunction Structures and Syntactic Domains." in KOLB, Hans-Peter & Uwe MÖNNICH (eds.): 259-299.
KRACHT, Marcus (2001): "Syntax in Chains." *Linguistics and Philosophy 24*: 467-529.
LARSON, Richard (1988): "On the Double Object Construction." *Linguistic Inquiry 19*: 335-391.
LASNIK, Howard (1995): "Case and Expletives Revisited." *Linguistic Inquiry 26*: 615-633.
LASNIK, Howard (1999a): *Minimalist Analysis*." Oxford: Blackwell.
LASNIK, Howard (1999b): "Chains of Arguments." in EPSTEIN, Samuel & Norbert HORNSTEIN (eds.): 189-215.
LASNIK, Howard (1999c): "On Feature Strength." *Linguistic Inquiry 30.2*: 197-217.

LASNIK, Howard & Joseph KUPIN (1977): “A Restrictive Theory of Transformational Grammar.” *Theoretical Linguistics 4.3*: 173-196.

LASNIK, Howard & Mamoru SAITO (1992): *Move-α*. Cambridge, Mass.: MIT Press.

LASNIK, Howard & Juan URIAGEREKA (1988): *A Course in GB Syntax*. Cambridge, Mass.: MIT Press.

LEBEAUX, David (1991): “Relative Clauses, Licensing, and the Nature of the Derivation.“ in *Syntax and Semantics 25*. New York: Academic Press. 209-239

LEUNINGER, Helen (1979): *Reflektionen über die Universalgrammatik*. Frankfurt/M.: Suhrkamp.

LUTZ, Uli, Gereon MÜLLER & Arnim von STECHOW (eds.) (2000): *WH-Scope Marking*. Amsterdam: John Benjamins.

MACDONALD, Cynthia (1995): “Introduction.“ in MACDONALD, Cynthia & Graham MACDONALD (eds.): 3.27.

MACDONALD, Cynthia & Graham MACDONALD (eds.) (1995): *Connectionism*. Oxford: Blackwell.

MANNA, Zohar & Richard WALDINGER (1985): *The Logical Basis for Computer Programming Vol.1*. Reading Mass.: Addison-Wesley.

MANASTER-RAMER, Alexis & Michael KAC (1990): “The Concept of Phrase Structure.“ *Linguistics and Philosophy 13*: 325-362.

MARANTZ, Alec (1995): “The Minimalist Program.“ in WEBELHUTH, Gert (ed.): 349-382.

McCAWLEY, James (1968): “Concerning the Base Component of a Transformational Grammar.“ *Foundations of Language 4*: 243-269.

McCAWLEY, James (1981): “The Syntax and Semantics of English Relative Clauses.“ *Lingua 53*: 99-149.

McCAWLEY, James (1982): “Parentheticals and Discontinuous Constituent Structure.“ *Linguistic Inquiry 13.1*: 91-106.

McCLOSKEY, James (1979): *Transformational Syntax and Model Theoretic Semantics*. Dordrecht: Reidel.

MICHAELIS, Jens (2001): *On Formal Properties of Minimalist Grammars*. Potsdam. [LiP 13].

MÜLLER, Gereon (1995): *A-bar Syntax*. Berlin: Mouton de Gruyter.

MÜLLER, Gereon (1998): *Incomplete Category Fronting*. Dordrecht: Kluwer.

MÜLLER, Gereon & Wolfgang STERNEFELD (eds.) (2001): *Competition in Syntax*. Berlin: de Gruyter.

MUSAN, Renate (1996): “'There' Constructions Revisited.“ *SALT VI*. 167-184.

MUYSKEN, Pieter (1982): “Parametrizing the Notion 'Head'“. *Journal of Linguistic Research 2*: 57-75.

MUYSKEN, Pieter & Henk van RIEMSDIJK (1986): “Projecting Features and Featuring Projections.“ in MUYSKEN, Pieter & Henk van RIEMSDIJK (eds.): *Features and Projections*. Dordrecht: Foris. 1-30.

NAKAMURA, Masanori (1998): “Reference Set, Minimal Link Condition, and Parameterization.“ in BARBOSA, Pilar et al. (eds.): 291-313.

NUNBERG, Geoffrey, Ivan SAG & Thomas WASOW (1994): “Idioms.“ *Language 70.3*: 491-538.

NUNES, Jairo (1995): *The Copy Theory of Movement and Linearization of Chains in the Minimalist Program*. PhD. Dissertation. University of Maryland.

NUNES, Jairo & Ellen THOMPSON (1998): “Appendix.“ in URIAGEREKA, Juan: *Rhyme and Reason*. Cambridge, Mass.: MIT Press.

NUNES, Jairo (1999): “Linearization of Chains and Phonetic Realization of Chain Links.“ in EPSTEIN, Samuel & Norbert HORNSTEIN (eds.): 217-249.

NUNES, Jairo (2001): “Sideward Movement.“ *Linguistic Inquiry 32.2*: 303-344.

OJEDA, Almerindo E. (1988): "A Linear Precedence Account of Cross-Serial Dependencies." *Linguistics and Philosophy 11*: 457-492.
OUHALLA, Jamal (1991): *Functional Categories and Parametric Variation*. London: Routledge.
OUHALLA, Jamal (1992): "Functional Categories, Agrammatism, and Language Acquisition." *Sprachwissenschaft in Frankfurt. Arbeitspapier Nr.9*.
PARTEE, Barbara (1975): "Montague Grammar and Transformational Grammar." *Linguistic Inquiry 6.2*: 203-300.
PARTEE, Barbara, TER MEULEN, Alice & Robert WALL (1993): *Mathematical Methods in Linguistics*. Dordrecht: Kluwer.
PEREIRA, Fernando (1986): "A Structure-Sharing Representation for Unification Based Grammar Formalisms." in SHIEBER, Stuart et al.: Part II, chapter 2. 17-35.
PESETSKY, David (2000): *Phrasal Movement and its Kin*. Cambridge, Mass.: MIT Press.
PETERS, Stanley (1972): "The Projection Problem." in PETERS, Stanley (ed.): 171-188.
PETERS, Stanley (ed.) (1972): *Goals of Linguistic Theory*. Englewood Cliffs: Prentice Hall.
PETERS, Stanley & Robert W. RITCHIE (1973): "On the Generative Power of Transformational Grammars". *Information Science 6*: 49-83.
PETERS, Stanley & Robert W. RITCHIE (1981): "Phrase-linking Grammar." Draft Ms. University of Texas at Austin.
POLLARD, Carl & Ivan A. SAG (1987): *Information-Based Syntax and Semantics*. Stanford: CSLI Publications.
POLLARD, Carl & Ivan A. SAG (1994): *Head-Driven Phrase Structure Grammar*. Chicago: Unversity of Chicago Press.
POLLOCK, Jean-Yves (1989): "Verb Movement, Universal Grammar, and the Structure of IP." *Linguistic Inquiry 20*: 365-424.
POLLOCK, Jean-Yves (1993): "Notes on Clause Structure." Ms. University of Picardie.
PULLUM, Geoffrey K. (1983): "How Many Possible Human Languages Are There?." *Linguistic Inquiry 14.3*: 447-467.
PULLUM, Geoffrey K. (1989): "Formal Linguistics Meets the Boojum." *Natural Language and Linguistic Theory 7*: 137-143.
PULLUM, Geoffrey K. (1996): "Nostalgic Views from Building 20." *Journal of Linguistics 32*: 137-147.
PUTNAM, Hilary (1975): "The 'Innateness Hypothesis' and Explanatory Models in Linguistics." in *Philosophical Papers Vol. 2*. Cambridge: CUP. 107-116.
QUINE, Willard V.O. (1959): *Methods of Logic*. New York: Holt, Rinehart, and Winston.
QUINE, Willard V.O. (1995/1941[1]): "Whitehead and the Rise of Modern Logic." in *Selected Logic Papers*. Cambridge, Mass.: Harvard University Press. 3-36.
REINHART, Tanya (1991): "Pronouns." in STECHOW, Arnim von & Dieter WUNDERLICH (eds.): *Semantik. Ein Internationales Handbuch der Zeitgenössischen Forschung*. Berlin: de Gruyter. 535-548.
REINHART, Tanya (1995): "Interface Strategies." *OTS-Working Papers 95-002*. Utrecht University.
REINHART, Tanya (1997): "Interface Economy and Markedness." in WILDER et al. (eds.): 146-169.
REULAND, Eric (1988): "Phrase Structure and the Theory of Levels". in CARDINALETTI, Anna, Guglielmo CINQUE & Giuliana GIUSTI (eds.): *Constituent Structure*. Dordrecht: Foris.
RICHARDS, Norvin (1999): "Featural Cyclicity and the Ordering of Multiple Specifiers." in EPSTEIN, Samuel & Norbert HORNSTEIN (eds.): 127-158.
RIEMSDIJK, Henk van & Edwin WILLIAMS (1986): *Introduction to the Theory of Grammar*. Cambridge, Mass.: MIT Press.

RIZZI, Luigi (1986): "On Chain Formation." *Syntax and Semantics 19*. New York: Academic Press. 65-95.
RIZZI, Luigi (1990): *Relativized Minimality*. Cambridge, Mass.: MIT Press.
ROBERTS, Ian (1998): "*Have/Be* Raising, Move F, and Procrastinate." *Linguistic Inquiry 29*: 113-125.
ROGERS, James (1998): *A Descriptive Approach to Language-Theoretic Complexity*. Stanford: CSLI-Publications.
ROORYK, Johan (1994): "Generalized Transformations and the Wh-Cycle." *GAGL 37*: 195-208.
ROSS, John R. (1986): *Infinite Syntax!* Norwood: Ablex.
RUSSELL, Bertrand (1993/1919[1]): *Introduction to Mathematical Philosophy*. London: Routledge.
SABEL, Joachim (1996): *Restrukturierung und Lokalität*. Berlin: Akademie-Verlag.
SAG, Ivan A. & Janet D. FODOR (1994): "Extraction Without Traces." *WCCFL XIII*: 365-384.
SAITO, Mamoru & Naoki FUKUI (1998): "Order in Phrase Structure and Movement." *Linguistic Inquiry 29.3*: 439-474.
SAUSSURE, Ferdinand de (1969/1915[1]): *Cours de Linguistique Générale*. Paris: Payot.
SCHOEMAKER, Paul J.H. (1991): "The Quest for Optimality: A Positive Heuristic of Science?" *Behavioral and Brain Sciences 14*: 205-245
SCHWARZSCHILD, Roger (1999): "Givenness, AvoidF, and other Constraints on Accent Placement." *Natural Language Semantics 7.2*: 141-177.
SHASTRI, Lokendra & Venkat AJJANAGADDE (1993): "From Simple Associations to Systematic Reasoning." *Behavioral and Brain Sciences 16*: 417-494.
SHIEBER, Stuart (1986): *An Introduction to Unification-Based Approaches to Grammar*. Stanford: CSLI-Publications.
SHIEBER, Stuart, Fernando PEREIRA, Lauri KARTTUNEN & Martin KAY (1986): "A Compilation of Papers on Unification-Based Grammar Formalisms Parts I and II." *Report No. CSLI-86-84*. Stanford.
SMULLYAN, Arthur F. (1971): "Modality and Description." in LINSKY, Leonard (ed.): *Reference and Modality*. Oxford: OUP. 35-43.
SPEAS, Margaret (1990): *Phrase Structure*. Dordrecht: Kluwer.
SPEAS, Margaret (1991): "Generalized Transformations and the D-Structure Position of Adjuncts." *Syntax and Semantics 25*. New York: Academic Press. 241-257.
SPERBER, Dan & Deirdre WILSON (1986): *Relevance*. Oxford: Blackwell.
STABLER, Edward P. (1983): "How are Grammars Represented?" *Behavioral and Brain Sciences 6*: 391-421.
STABLER, Edward P. (1992): *The Logical Approach to Syntax*. Cambridge, Mass.: MIT Press.
STABLER, Edward P. (1996): "Acquiring and Parsing Languages With Movement." Ms. UCLA.
STABLER, Edward P. (1998): "Acquiring Languages With Movement." *Syntax 1*: 72-97.
STAUDACHER, Peter (2000): "Partial Movement and Compositionality." in BAYER, Josef & Christine RÖMER (eds.): *Von der Philologie zur Grammatiktheorie.* Tübingen: Niemeyer. 191-211.
STECHOW, Arnim von (1993): "Die Aufgaben der Syntax." in JACOBS, Joachim et al. (eds.): 1-88.
STECHOW, Arnim von & Wolfgang STERNEFELD (1988): *Bausteine Syntaktischen Wissens*. Opladen: Westdeutscher Verlag.
STEGMÜLLER, Wolfgang (1987): *Hauptströmungen der Gegenwartsphilosophie Vols. I+II*. München: Kröner.
STERNEFELD, Wolfgang (1991): *Syntaktische Grenzen*. Opladen: Westdeutscher Verlag.
STERNEFELD, Wolfgang (1997): "Comparing Reference Sets." in WILDER et al. (eds.): 81-114.
SUPPES, Patrick (1972): *Axiomatic Set Theory*. New York: Dover.

TOYOSHIMA, Takashi (1997): “Head-to-Spec Movement.“ Ms. Cornell University.
TRUDEAU, Richard (1993): *Graph Theory*. New York: Dover.
URIAGEREKA, Juan (1998): *Rhyme and Reason*. Cambridge, Mass.: MIT Press.
URIAGEREKA, Juan (1999): “Multiple Spell-Out.“ in EPSTEIN, Samuel & Norbert HORNSTEIN (eds.): 251-282.
VOGEL, Ralf & Markus STEINBACH (1994): “Zum Konzept der Tiefenstruktur in der Generativen Grammatik.“ M.A. Thesis. Frankfurt/M. University.
WARTENA, Christian (1994): *Spelling Out the Minimalist Program*. M.A.Thesis. University of Nijmegen.
WARTENA, Christian (2000): *Storage Structures and Conditions on Movement in Natural Language Syntax*. Ph.D. Dissertation. University of Potsdam.
WATANABE, Akiro (1995): “Conceptual Basis of Cyclicity.“ *MITWPL 27*: 269-291.
WEBELHUTH, Gert (1995): “X-Bar Theory and Case Theory.“ in WEBELHUTH, Gert (ed.): 15-95.
WEBELHUTH, Gert (ed.) (1995): *Government and Binding Theory and the Minimalist Program*. Oxford: Blackwell.
WIESE, Richard (1996): “Phrasal Compounds and the Theory of Word Syntax.“ *Linguistic Inquiry 27.1*: 183-193.
WILDER, Chris (1991): “Tough Movement Constructions.“ *Linguistische Berichte 132*: 115-132.
WILDER, Chris (1995): “Antecedent Containment and Ellipsis.“ in ALEXIADOU, Artemis et al. (eds.): *FAS Papers in Linguistics 4*. 132-165.
WILDER, Chris (1996): “Rightward Movement as Leftward Deletion.“ in LUTZ, Uli & Jürgen PAFEL (eds.): *On Extraction and Extraposition in German*. Amsterdam: John Benjamins. 273-309.
WILDER, Chris (1997): “Phrasal Movement in LF.“ *NELS 27:* 425-439.
WILDER, Chris & Damir ĆAVAR (1994): “Word Order Variation, Verb Movement, and Economy Principles.“ *Studia Linguistica 48*: 46-86.
WILDER, Chris & Hans-Martin GÄRTNER (1997): “Introduction.“ in WILDER et al. (eds.): 1-35.
WILDER, Chris, Hans-Martin GÄRTNER & Manfred BIERWISCH (eds.) (1997): *The Role of Economy Principles in Linguistic Theory*. Berlin: Akademie-Verlag.
WILLIAMS, Edwin S. (1989): “Maximal Projections in Words and Phrases.“ in BALTIN, Mark & Anthony KROCH (eds.): 280-291.
WINOGRAD, Terry (1987): “Thinking Machines.“ *Report No. CSLI-87-100*. Stanford.
ZWART, Jan-Wouter (1997): *Morphosyntax of Verb Movement*. Dordrecht: Kluwer.
ZWART, Jan-Wouter (2001): “Syntactic and Phonological Verb Movement.“*Syntax 4.1*: 34 - 62
ZWARTS, Joost (1992): *X'-Bar-Syntax-X'-Semantics*. Utrecht: OTS Dissertation Series.