Introduction

Samuel David Epstein, Höskuldur Thráinsson, C. Jan-Wouter Zwart

This introduction has two parts. The first part is a general introduction to the Minimalist Program. It opens with a section on the common goals of modern linguistic theory, explaining how the Minimalist Program addresses questions that have guided generative research from its beginnings (section 1.1).The second part is an introduction to the articles in this volume. Apart from the opening section of part one, the two parts are structured alike. They address the following issues: phrase structure theory (section 1.2 and 2.1), movement and syntactic licensing (1.3 and 2.2), economy conditions (1.4 and 2.3), the properties of the structures built and the typology of syntactic positions (section 1.5 and 2.4), parametric variation (1.6 and 2.5), and directionality and word order (1.7 and 2.6). The reader will want to check part 1 of the introduction for a quick survey of the central notions of the Minimalist Program. Part 2 adds critical discussion, puts these notions in (historical) context, and makes the various contributions of the articles in this volume explicit.

1. Introduction to the Minimalist Program

1.1 Origin and some basic properties

1.1.1 Common goals

Two of the central goals of linguistic theory are:

(1) a. To provide an explicit **description** of what constitutes knowledge of each natural language.
b. To provide an **explanation** of how it is that a person P can come to know what they do know.

(1a) implies that for every possible natural language we must answer the question "What does a person P, who knows a language L, know?". (1b) means that for each natural language L that is humanly knowable we must explain how it is that a person P can come to know L. More specifically, (1b) requires that we answer two questions:

(2) a. What is the contribution of the organism?
b. What is the contribution of the environment?

As Chomsky (1995a:387) notes:

> The first efforts to address these problems quickly revealed that traditional grammatical and lexical studies do not begin to describe, let alone explain, the most elementary facts about even the best-studied languages ... This is hardly a discovery unique to linguistics. Typically, when questions are more sharply formulated, it is learned that even elementary phenomena had escaped notice, and that intuitive accounts that seemed simple and persuasive are entirely inadequate. If we are satisfied

> that an apple falls to the ground because that is its natural place, there will be no serious science of mechanics. The same is true if one is satisfied with traditional rules for forming questions, or with the lexical entries in the most elaborate dictionaries, none of which come close to describing simple properties of these linguistic objects.
>
> Recognition of the unsuspected richness and complexity of the phenomena of language created a tension between the goals of descriptive and explanatory adequacy. It was clear that to achieve explanatory adequacy, a theory of the initial state must hold that particular languages are largely known in advance of experience. The options permitted in universal grammar (UG) must be highly restricted; limited experience must suffice to fix them one way or another, yielding a state of the language faculty that determines the varied and complex array of expressions, their sound and meaning, in a uniform and language-independent way. But this goal receded still further into the distance as generative systems were enriched in pursuit of descriptive adequacy, in radically different ways for different languages. The problem was exacerbated by the huge range of phenomena discovered when attempts were made to formulate actual rule systems.

These rule systems formulated took each grammar/language to be

> a rich and intricate system of rules that are, typically, construction-particular and language-particular: the rules forming verb-phrases or passives or relative clauses in English, for example, are specific to *these* constructions in *this* language. (Chomsky 1993:4)

Thus, there arose the essential tension common to all scientific inquiry between descriptive adequacy (empirical coverage) on the one hand and explanatory adequacy on the other, with early theories arguably erring by virtue of being overly descriptive, incorporating myriad language-specific and even construction-specific postulates. Thus,

> In essence, the problem is that there are too many rule systems. Therefore it is hard to explain how children unerringly select one such system rather than another. (Chomsky 1987)

Within the more recent Principles and Parameters approach, it was explicitly recognized that these very serious obstacles to explanatory adequacy are created by the incorporation of construction-specific and language-specific theoretical constructs. An explanatory theory of language, not a theory incorporating postulates specific to English passives or French existential sentences, i.e. specific to the data, was sought. This theory, the theory of universal grammar, was assumed to consist of principles and parameters. The former are fixed and invariant language-independent, construction-independent universal linguistic laws common to all human languages. The parameters, invoked to capture seemingly irreducible and real differences among languages are the vestiges of the language particular rules of the earlier standard theory. There is assumed to be a finite (if not very small) number of parameters, each with a finite number of (perhaps two) values. Thus cross-linguistic variation, as expressed by parameters, is assumed to be highly restricted. A recent even more circumscribed theory of the formal properties of parameters postulates that each parameter is expressible as an irreducible cross-linguistic difference in the value of (a circumscribed set of) morphological features associated with certain lexical categories—the so-called functional syntactic categories such as Inflection/Agreement, Complementizer, and Determiner.

As Chomsky notes (1995a:387) this theory of Principles and Parameters is consonant with Jespersen's view that it is only "with regard to syntax" that we expect that "there must be something in common to all human speech"—a "universal (or general) grammar", though "no one ever dreamed of a universal morphology".

If the Principles and Parameters theory is on the right track, then "...there is only one computational system [syntax] and one lexicon, apart from this limited [morpho-parametric] kind of variety" (Chomsky 1993:3)

Thus a unified explanatory theory was sought—counter-intuitively postulating that there is, in effect, only one human language despite the appearance of unlimited diversity. The Minimalist Program—a research program—represents the most recent and direct outgrowth of this ongoing inquiry.

Perhaps the most distinctive feature of this Program—and that which makes it most exciting and promising—is its explicit programmatic commitment to explanation through the characteristic method of minimization, consistent with Einstein's (1954) perspective that "... the grand aim of all science is to cover the greatest possible number of experimental facts by logical deductions from the smallest number of hypotheses or axioms." As concerns this minimization of the axioms, Nash (1963:173) rightly notes that "[e]ach 'quality' imputed to a premised entity figures as an additional postulate. Our desire for parsimony of postulates thus evokes a search for theoretical posits having the slenderest possible qualitative endowment".

Consistent with this search for that which we might call "simplicity", Chomsky presents the working hypothesis that

> ... languages are based on simple principles that interact to form often intricate structures... the language faculty is non-redundant, in that particular phenomena are not "overdetermined" by the principles of language (1993:2) ... the principles outlined are simple and restrictive... (1993:5)

Once committed to "explanation through minimization" there of course emerge, among others, the entirely formal and empirical questions: "Which aspects of the existing theory should be simplified, which eliminated, and which retained (if any)?" The answers, regardless of the domain of inquiry, are never self-evident, nor the totality of their empirical consequences immediately (if ever) determinable.

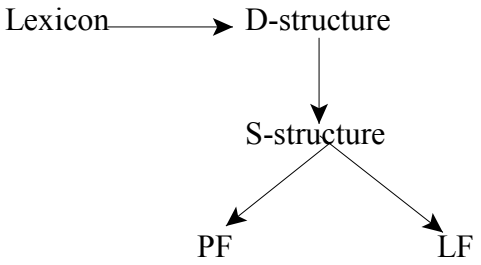
The Minimalist Program is the most recent (programmatic) outgrowth of the theory of generative grammar originally developed in Chomsky (1957, 1965). The immediate predecessor of the Minimalist Program—the so-called Government and Binding Theory (Chomsky 1981, 1982, 1986a, b)—upon informed, studied and insightful scrutiny was argued to exhibit certain suspect formal properties. The Minimalist Program first identifies, through careful and insightful analyses, potentially problematic aspects of the so-called GB theory and then proceeds to explore alternative conceptions of grammar lacking these arguably unattractive properties, eliminating them altogether or replacing them with more natural, less stipulative constraints.

1.1.2 Minimizing Levels of Representation

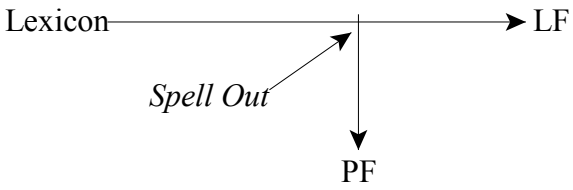
Perhaps the first targets of minimization were levels of representation. Unlike the four-level Government and Binding model of Chomsky (1981), illustrated in (3a), the Minimalist

Program assumes only two levels of representation—one an abstract representation of sound (PF), the other an abstract representation of meaning (LF), each seemingly ineliminable (3b):

(3) a.



b.



The levels of D-structure and S-structure in (3a), which are internal representations, in the sense that they feed into no system external to the syntactic component, are eliminated.

In the absence of a level of D-structure representation, there nonetheless remains a Lexicon and two concatenative and formally simple binary transformational (structure building) operations (cf. (3b)). One is Merge, a resurrection of Generalized Transformations postulated in Chomsky (1957) and later abandoned in favor of a model in which recursion is performed by (non-transformational) phrase-structure rule application. The other concatenative operation is Move. Each of these operations concatenates two and only two categories, forming a third. The rules apply iteratively (subject to the strict cycle condition (Chomsky 1973), reformulated as an ``extension condition" (Chomsky 1993:23)), thereby building constituent structure.

At any point in the process of such iterative binary concatenative rule-application, a (non-transformational) rule/operation called Spell-Out can be optionally applied. The result of such rule application is that from the structure thus far derived certain information is split off and sent to the phonological component, while the derivation continues after the split off point continues, yielding the logical form (or covert) component. Thus, the split between PF and LF components inherent in the Government and Binding model is retained—the empirical hypothesis that neither component interacts with the other remains. However, D-structure is abandoned, an attractive and seemingly viable hypothesis especially given an articulated (copy) theory of traces within which all relevant aspects of prior representation are apparently retainable throughout the derivation (cf. Freidin 1978, Koster 1987).

As Chomsky (1993:21) notes, recent research, including Reinhart (1991), Kroch and Joshi (1985), Kroch (1989), Lebeaux (1988), and Epstein (1991) also raises empirical problems engendered by the postulation of a D-structure level of representation in which all syntactic categories appearing in a derivation D must appear. In addition, Sportiche (1983), Epstein (1987), Larson (1988), and research concerning *tough*-constructions (in particular Lasnik and Uriagereka 1988:147 and Brody 1993) also present analyses inconsistent with either the all-

at-once property of D-structure (to use Chomsky's term) and/or inconsistent with a defining property of D-structure, namely the application of the Theta Criterion at this level.[1] Much of this research relies, in place of D-structure, upon Generalized Transformations, with lexical access/insertion consequently applying over the course of the derivation–although, as Chomsky (1993: 21) notes "... the empirical consequences of the D-Structure conditions remain to be faced... ."

S-structure as an independent level of representation is, as noted above, also eliminated, being replaced by the optional rule of Spell-Out, allowed to apply at any point—splitting the derivation into two tracks exactly as in the Government and Binding-model. This elimination of S-structure was already suggested in Chomsky (1986a) wherein it is observed that S-structure has a unique status within the Government and Binding model in that it is the sole (central) level linking, i.e. having contact with each of the other three levels—D-structure, PF, and LF. As Chomsky observed, this raises the possibility that its properties may well be deducible from the three requirements that: (i) it must be derivable from D-structure by iterative transformational-rule application and, at the same time (ii) PF representation must be derivable from it by the operations internal to the PF component, and (iii) LF representation must be derivable from it by iterative transformational-(covert)-rule application. As Chomsky (1986a:101) notes:

> properties of S-structure may be reducible to the independent conditions of Full Interpretation holding of PF and LF representation ... given an appropriate account of the ways in which the elements of a structure Σ [i.e. the levels of representation E/T/Z] may be related.

It is this "reduction to independent conditions" that pervades the Minimalist Program. As concerns S-structure representation, the idea is that it is eliminable by appeal to an optional rule of Spell-Out which is allowed to apply at any point in the derivation.

The hypothesis is that this level elimination and its concomitant replacement with an optional, unordered and virtually conceptually necessary operation (ultimately yielding two ineliminable representations: one of sound, the other of meaning) may well be empirically adequate. The precise derivational point at which Spell-Out applies is, by hypothesis, deducible from independent requirements (Full-Interpretation) imposed on the only two (and the natural two) levels of representation: PF and LF.

To take but one illustrative example, consider the representations in (4):

(4) a. [*was arrested he*]
b. [*he was arrested*]

Were Spell-Out to apply to the representation in (4a), the claim is that this is not a licit sound, specifically, the Case-requirements imposed upon *he*, requirements now interpreted as morphophonetic PF-requirements (S-structure having been eliminated) are not satisfied. If, on the other hand, Spell-Out were applied after NP-movement of *he* to subject position, yielding (4b), this representation, submitted to PF scrutiny, is regarded as consisting entirely of legitimate, i.e. legible, PF objects. In particular *he* is PF-licit by virtue of satisfying Case-

[1] Notice that this defining property is conceptually problematic as well: the Theta Criterion, if it exists, is an interpretive condition, and should therefore apply only at LF.

requirements, requirements which always have been construed as pertaining to (at least) phonetically realized NPs/DPs (cf. Chomsky 1981:49).

Thus, it is hypothesized that there is neither a D-structure- nor an S-structure-level of representation but rather a Lexicon, a quite simple binary concatenative algorithm yielding, for each expression, a PF (sound) representation and an LF (meaning) representation.

As expected, the properties of these two remaining levels of representation—PF and LF—are themselves also subjected to scrutiny. Thus one must ask the following questions:

(5) a. Why are there two levels of representation (i.e. PF and LF)?
b. Why do they each have the properties they seem to have—i.e. Why are PF and LF representations each constituted of certain objects (not others), and why must the objects be arranged in a particular way?

Chomsky conjectures that the answers lie in the fact that representations at PF and LF (the interface levels) must be read by (legible to) the Articulatory-Perceptual (A-P) and the Conceptual-Intentional (C-I) systems respectively, that is—the A-P and C-I systems external to the syntactic component proper, impose legibility conditions on the interface representations.

For example, a PF representation must, by hypothesis, be linearized in order for it (or some transform of it) to undergo pronunciation, i.e. interpretation by the A-P system. Thus, a fairly fundamental aspect of PF representations—that they contain linearly ordered strings—is arguably deducible from demands imposed from without, i.e. requirements imposed by the articulatory system. Similarly, the following entirely descriptive universal (substantive) filter

(6) *V
+high
+low

may in fact not be a descriptive phonological filter, but might instead be explained by appeal to natural properties of the articulatory systems structurally external to the syntactic component but which nonetheless take as input, interface representations (in this case, PF representations).

These legibility conditions imposed on PF and LF representations by the A-P and C-I systems respectively are known as the Bare-Output Conditions and here again the program looks promising and, importantly, the questions it seeks to address must be posed in any event.

If on the right track—the number and properties of the only two and the natural two representations—those of sound and meaning are to be deduced from the properties of the external systems, A-P and C-I which (in both senses of the word) must read these representations. Thus the Minimalist Program seeks to eliminate, on principled grounds, D-structure representation and S-structure representation while concomitantly hypothesizing a programmatic deduction of the properties of the only two (comparatively uncontroversial and seemingly ineliminable) levels of representations—the interface levels of PF and LF represention.

1.1.3 Building Representations

Having discussed the Minimalist reduction of levels of representation, we now must address the question of how representations (PF and LF representations) are generated.

As noted, there is assumed to be a Lexicon—an arguably irreducible component of the grammar expressing what we know when we know the words of a given language. Each lexical entry is assumed to consist of, at least, a set of three sets: a semantic-feature set, a phonological-feature set, and a syntactic-feature set. Again this seems unavoidable—a PF representation containing *cat* /kæt/ must provide instructions to the A-P system specifying that this particular lexical item (meaning what it does) is pronounced /kæt/, not /tæk/ nor /ækt/. Thus a lexical entry must contain irreducible/unpredictable phonological specifications to be interpreted when PF representations are submitted to the A-P system. The same holds of semantic properties of lexical items—these are tantamount to irreducible atomic instructions (present in LF representations containing lexical items) to the C-I system ("interpret *cat* as ..."). Importantly then, it follows that certain aspects of, or constraints on, lexical representations are deducible from the bare-output conditions—each aspect of a lexical representation (present in a PF or in an LF representation) must be interpretable by A-P and C-I systems, respectively. Hence we have a program for determining the core notion "possible lexical representation".

In addition to phonological features and semantic features a lexical item is assumed to have syntactic features, e.g. categorial features (Noun) and φ-features (including person, number, and gender) identifying its syntactic status and providing instructions to the syntactic combinatorial system itself.

Given a Lexicon, the generative procedure consists of Merge and Move (see 1.2 below). Each is subject to the strict cycle condition (Chomsky 1973) and each concatenates exactly two objects forming a third. For purposes of illustration, we might have a derivation such as:

(7) a. 1st *Merge* the_{D} *and* dog_{N} *yielding* $[_{DP}\ \text{the}_{D}\ \text{dog}_{N}]$

b. 2nd *Merge* arrested_{V} *and* $[_{DP}$ the dog$]$ *yielding* $[_{VP}\ [_{V}\ \text{arrested}]\ [_{DP}\ \text{the}_{D}\ \text{dog}_{N}]]$

c. 3rd *Merge* be_{Infl} *and* $[_{VP}\ [_{V}\ \text{arrested}]\ [_{DP}\ \text{the}_{D}\ \text{dog}_{N}]]$ *yielding*
$[_{I'}\ \text{be}\ [_{VP}\ [_{V}\ \text{arrested}]\ [_{DP}\ \text{the}_{D}\ \text{dog}_{N}]]]$

d. 4th *Merge* $[_{DP}$ the dog$]$ *with* $[_{I'}\ \text{be}\ [_{VP}\ [_{V}\ \text{arrested}]\ [_{DP\text{TRACE}}\ \text{the}_{D}\ \text{dog}_{N}]]]$
yielding
$[_{IP}\ [_{DP}$ the dog$]\ [_{I'}\ \text{be}\ [_{VP}\ [_{V}\ \text{arrested}]\ [_{DP\text{TRACE}}\ \text{the}_{D}\ \text{dog}_{N}]]]]$

Ultimately [the dog **was** arrested $[_{DP\text{TRACE}}\ \text{the}_{D}\ \text{dog}_{N}]]]]$
(via application of agreement)

The 1st Merge concatenates two lexical items. The 2nd and 3rd Merge each concatenates one lexical (atomic) item and an already constructed (molecular) phrase. The 4th operation, involving Move, takes a phrase within an already constructed tree and concatenates it with that tree. This too is binary—concatenating one member of a tree with the one tree containing it. The strict cycle condition (Chomsky 1973) prohibits (syntactic) infixation. The strict cycle condition, now called the ``extension condition", dictates that no category can be concatenated with a category (already) having a dominating node. Informally, Merge and

Move always add a sister to the root node.[2]

In the following sections (sections 1.2-1.6) we will introduce the basic concepts of the Minimalist Program in more detail. In section 1.7, we will present a brief sketch of the antisymmetry hypothesis of Kayne (1994), which complements the Minimalist Program outlined in Chomsky (1993).

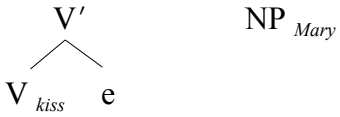
1.2 Building up phrase structure

As just described, representations are built up in a (strict-cyclic) bottom-up fashion by Generalized Transformation. A Generalized Transformation combines two phrase markers where here we include among the class of phrase markers, lexical items, e.g. [$_{N}$cat]. Two phrase markers are combined by expanding one (the target phrase marker) so as to include an empty position. This expansion takes place by adding to the target phrase marker a projection of the target phrase marker. This projection is binary branching and has two daughters: the target phrase marker and an empty position. The other phrase marker then substitutes into this empty position. The whole process, illustrated in (8), yields two sister phrase markers in a binary branching subtree.[3]

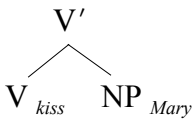
(8) a. Two independent phrase markers

V *kiss* NP *Mary*

b. Expansion of the target phrase marker V



c. Substitution of NP into the empty position in the projection of the target phrase marker



a-c apply internal to a single operation. There is no intermediate representation of the form

[2] This condition may itself be deducible from natural and independently motivated conditions (Kitahara 1994, 1995, 1996). Moreover, Merge and Move—each being binary concatenative operations may well be unifiable as a single operation Target-α as proposed by Kitahara (1994, 1995, 1996).

[3] Binary branching is a result of this particular formulation of the Generalized Transformation mechanism. The attractiveness of binary branching has been argued for several times in the literature (see e.g. Kayne 1984, E. Hoekstra 1991).

(8b). The projection of the target phrase marker has the same categorial features as the target phrase marker. The phrase level of the projection of the target phrase marker is determined by the X-bar schema (Chomsky 1986b, Chomsky 1970, Jackendoff 1977), specifying that the ultimate projection of an X (or X°, or head) will be an XP (or X″, or maximal projection), and that there is an intermediate projection X′ (X-bar) which is the immediate projection of X. This is expressed in the following two rewrite rules:

(9) a. XP → (ZP) X′
b. X′ → (YP) X°

The order of the elements to the right of the arrows in (9) is irrelevant (but see section 1.7 below). The sister of X°, YP in (9b), is called complement; the sister of X′, ZP in (9a), is called specifier.

Because the Generalized Transformation illustrated in (8) combines two independent phrase markers, it is called a binary operation. Lexical insertion is a typical binary operation. It is also possible that the empty element created by expanding the target phrase marker is substituted for by an element contained in the target phrase marker. This would be called a singulary operation, or Move, e.g. (7d) above.

1.3 Licensing: movement and feature checking

A standard distinction exists in linguistic theory between contentful elements and functional elements. Word stems are contentful elements, whereas inflectional morphemes are functional elements. Functional elements express agreement relations between constituents.

In the Minimalist Program, it is assumed that agreement relations are highly local. A maximal projection α agrees with a head β only if α is a specifier of β. Thus in (7) there is a Spec-Head relation between [the dog] (singular) and *be*: here singular overt morphological agreement (as well as Past Tense morphology) (*was*) appears.[4] A head α agrees with a head β only if α is adjoined to β.[5] Moreover, β must be a functional head.

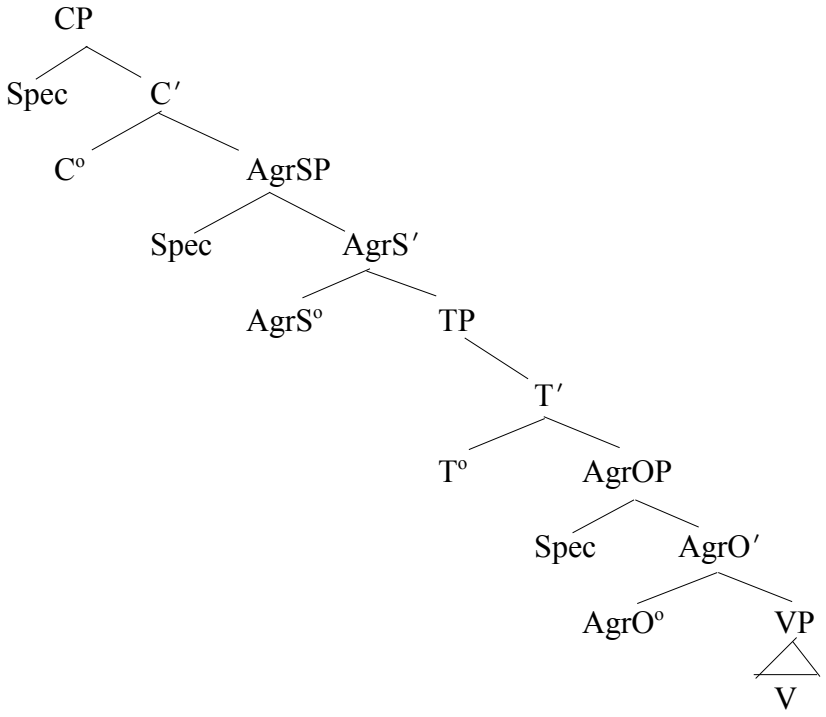
In the Government and Binding framework, the distinction between contentful (or lexical) elements and functional elements gradually took the following shape.[6] Functional elements are generated as heads of independent phrasal projections. This idea has roots going back to the Affix-Hopping analysis of Chomsky (1957, 1965) in which at D-structure agreement morphemes are generated separated from the verb to which they are affixed in the S-structure representation. Thus, for example, the inflectional morphemes tense and agreement are generated separate from the lexical stems. This yields a structure as illustrated in (10):

[4] The Spec-Head relation can be instantiated before and after Spell-Out. Possibly, in the English example *the dog* and *was* are not in a Spec-Head relation in overt syntax, since adverbs may intervene (*the dog probably was not arrested*).

[5] The locality requirements are further restricted in Zwart (1992b), where it is argued that an element α agrees with an element β only if α adjoins to β. This implies that in a specifier-head agreement relation, the specifier does not agree with the head, but with the immediate projection of a head. See also Epstein (1994, 1995).

[6] The developments in the Government and Binding period are marked by Stowell (1981) and Pesetsky (1982), on the structure of IP, Chomsky (1986b) on the structure of CP, Kayne (1989), Pollock (1989) and Chomsky (1991) on the division of IP into AgrP and TP, and Abney (1987) on the functional domain of noun phrases, DP. See Fukui and Speas (1986) on the relevance of functional projections for parametrization.

(10)



In (10), C° stands for the complementizer position, T° for tense, and AgrS° and AgrO° for subject and object agreement morphology, respectively. These functional heads project phrases in accordance with the X-bar schema given in (9) above. AgrOP, TP, AgrSP and CP together constitute the functional domain of a syntactic structure. VP constitutes the lexical domain. The stems, generated in V, have to be united with the inflectional morphemes in the functional heads through a transformational process of adjunction.

In the Minimalist Program, this analysis is maintained in a simplified form. The major difference concerns the content of the lexical and functional heads. In the Minimalist Program, lexical heads are fully inflected forms (stems plus inflectional affixes). These forms carry a feature associated with the inflectional affix. The functional heads likewise consist of features associated with inflectional morphology.[7]

The features associated with the inflectional morphology of lexical categories have to match the features represented in the functional heads. Matching is checked under the same strict locality requirements as is agreement (in fact, agreement is a subcase of feature matching). Thus again returning to (7d), there would be a feature mismatch appearing in Spec-Head relation in e.g. [*the dogs* (plural)] *was* (singular) *arrested*]. Therefore, the requirement that morphological features match triggers movement of lexical elements to

[7] The assumption that abstract features associated with inflectional morphology are of greater syntactic significance than the overt morphology itself is already a crucial part of the Case Theory module of the Government and Binding framework. This Case Theory refers to abstract Case features which are associated with nouns and noun phrases regardless of the morphological manifestation of Case on the nouns and noun phrases (Vergnaud 1982, Chomsky 1981). This theory of abstract Case is subsumed under the Minimalist Program. As a result, the inflectional features associated with Case are assumed to be present on lexical categories, even if there is no overt morphological manifestation of Case on these categories.

positions in the functional domain. Licensing inflected elements consists in moving the inflected elements to positions in the functional domain, and checking whether the features associated with the inflection match the features represented in the functional heads.

Recall that movement is an application of Generalized Transformation. The structure in (10), therefore, is completely built up in the process of moving elements from the lexical domain to positions in which their features can be checked (which yields the functional domain). There is no top-down phrase-structure rule system ensuring that syntactic structures are always like (10). The structure in (10) is the result of the fact that inflected elements must be licensed outside of the lexical domain.

The inflectional features relevant to the phenomena of verb movement and noun phrase movement include tense, agreement, and Case.[8] It is quite possible that other similarly relevant features exist, but these three appear to be indispensable.[9]

The features represented in the functional heads trigger both head movement (to the functional heads) and XP-movement (to the specifier positions of the functional heads). For this reason, Chomsky (1993) distinguishes two types of features borne by the functional heads: N-features and V-features. N-features are relevant for checking features of XPs (maximal projections), V-features are relevant for checking features of heads.

1.4 Economy conditions and restrictions on movement

In direct contrast to the Move-α hypothesis, transformational applications are subject to general conditions of economy. The derivation should take as few steps as possible (economy of derivation), and moreover the resulting representations should have as few symbols as possible (economy of representation) (Chomsky 1991). As Chomsky suggests (1993:5)

> ... it seems that economy principles of the kind explored in early work play a significant role in accounting for properties of language. With a proper formulation of such principles, it may be possible to move toward the minimalist design: a theory of language that takes a linguistic expression to be nothing other than a formal object that satisfies the interface conditions in the optimal way.

One aspect of economy of derivation is that movement always takes the shortest route. Shortness can be interpreted in two ways, viz. as involving the fewest number of rule

[8] The exact difference between Case and agreement is not very clear in this system. It is assumed that the specifiers of AgrS and AgrO are the positions for checking Nominative and Accusative Case features respectively. This suggests that Case and agreement are identical concepts. However, Chomsky (1993: 29-30) suggests that, while Nominative and Accusative Case features are checked in the specifier positions of AgrS and AgrO, respectively, the features relevant for checking Case do not reside in AgrS and AgrO, but in T and V, respectively (see 1.5). See also the discussion in the papers by FERGUSON and JONAS in the present volume.

[9] Iatridou (1990) contends that the approach to inflectional morphology sketched here leads to an "explosion" of functional categories, assuming that every functional category discovered in studying the languages of the world should be present in the grammar of every single language of the world. One way to think of this is to assume that a small number of inflectional features are present in all languages of the world, whereas a larger number may be relevant to specific languages only. What is syntactically universal, however, is the way the presence of inflectional features determines movement and word order. - For further discussion of issues of this kind see the papers by SOLÀ and THRÁINSSON in the present volume.

applications (Shortest Derivation) and/or as involving the shortest movements (also known as the Minimal Link Condition). These two interpretations appear to be contradictory (cf. Chomsky 1993:21). Another consequence is that any movement that is not triggered by a requirement of morphological feature checking is excluded. Thus, elements, once licensed, are immobile. This is known as Last Resort (see e.g. Epstein 1992 for recent discussion).

Economy of representation (naturally enough) excludes the presence of uninterpretable material in an interface representation. One instantiation of this is the principle of Full Interpretation (FI).[10]

FI requires that every element of an interface representation must provide a meaningful/legible input to the syntax-external cognitive systems, A-P and C-I. Only these elements are considered to be legitimate objects in an interface representation. Again such requirements (the Bare Output Conditions) seem natural assuming that interface representations are input to the A-P and C-I systems.

The syntactic features associated with inflectional morphology are considered to be relevant for syntax only, i.e. they lack, by hypothesis, phonological features and semantic features and therefore are uninterpretable at the interface levels. They play a crucial part in the licensing of inflected elements. However, these features are uninterpretable to components of the cognitive system external to the syntax. In other words, the features associated with inflectional morphology are not legitimate objects at the interface level: they cannot be a part of an interface representation that is to serve as input to other components of the cognitive system.

For this reason, these features must be eliminated during the derivation. It is assumed that matching features on a purely inflectional head are eliminated as soon as they are checked.[11]

Therefore, a minimal number of derivational steps is required to achieve a minimal representation at the interface of the syntactic component and other components of the cognitive system.

Two other economy principles are directly relevant.

First, picture the derivation as a step-wise procedure. At each step, economy of derivation will allow only a minimum of transformational activity. Eventually, movements will have to take place, but economy of derivation dictates that these take place as late in the derivation as possible. This can be formulated as a separate principle, Procrastinate (Chomsky 1993: 30).

Second, movement is triggered by the need to license inflected elements–more exactly, by the need to check off and thereby delete the uninterpretable abstract features associated with inflected elements, which, recall are prohibited at the interface. Elements that are already licensed, or that do not need licensing, are not forced, hence they are not allowed, to move. It follows that such elements can never move, not even in order to assist in the licensing of another element. The economy principle that prohibits such solely altruistic movement is called Greed (Chomsky 1993: 33).

[10] This principle was first introduced in Chomsky (1986a:98), in the context of a discussion of the relation between Case assignment and theta-role assignment. The idea was that noun phrases must be assigned Case at S-structure, because only then would they be visible for theta-role assignment at LF. Since only noun phrases that carry a theta-role are interpretable at the interface of LF and other components of the cognitive system, the principle of Full Interpretation requires Case assignment at S-structure. This concept has been slightly changed in the Minimalist Program. Case checking eliminates features since Case cannot be interpreted at the interfaces. Without Case checking, Full Interpretation (and economy of representation) is violated, since unchecked features are uninterpretable.

[11] This is the current incarnation of recoverable “deletion-under-identity”.

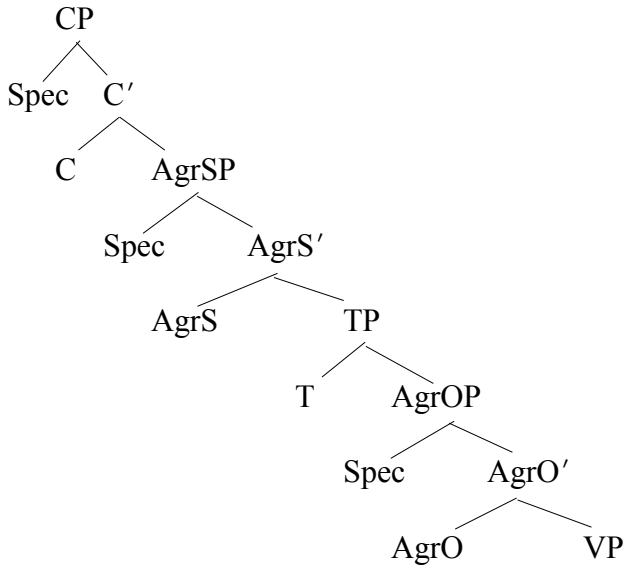
1.5 Phrase structure and category types

Crucially, rule application is highly restricted within the minimalist system. This stands in direct contrast to the Move-α hypothesis, under which transformational rule application was entirely unconstrained. D-structure is eliminated and representations are built from the bottom up, step by step (i.e. in a succession of binary rule applications), subject to the strict cycle/extension condition, with Move further restricted/defined by (at least) the derivational economy constraints, Greed and Procrastinate.

The syntactic component is thus, in Chomsky's terms, strongly derivational. Intermediate representations and the formal properties of the recursive procedure generating them are syntactically significant–an empirical hypothesis antithetical to the entirely filter-based (representation-characterizing) Move-α hypothesis.[12] In the Move-α approach, the generative procedure itself is unconstrained (if not undefined). Properties of the output representations (however generated) are the sole object of linguistic significance.

As noted, Chomsky (1993), adapting Pollock's (1989) analysis, assumes the following basic clause structure:[13]

(11) (= Chomsky 1993: example (2), p. 7 (slightly modified))



[12] For further exploration of this derivational approach see Epstein (1993, 1994, 1995), Epstein, Groat, Kawashima, and Kitahara (forthcoming), Groat (1995, in progress), and Kitahara and Kawashima (1994).

[13] As Chomsky (1993: 7) notes, this representation omits possible additional structure. As regards the position Spec-TP, see JONAS (1992, 1994, 1995, this volume), Jonas and Bobaljik (1993), Bobaljik and Jonas (1993), Bures (1992, 1993), Bobaljik and Carnie (1992), and THRÁINSSON (1993a, this volume). Also omitted is a possible Negation and/or Affirmation Phrasal Projection (see Pollock 1989, Laka 1990, Zanuttini 1991).

Here INFL of Chomsky (1981) is split into two distinct X-bar-projections – a Subject-Agreement (AgrS) projection and a Tense (T) projection. In addition, an object-agreement (AgrO) phrase appears as sister of T˚, and its head AgrO takes VP as its complement. Agr is assumed to be a collection of φ-features (person, number, and gender) playing a central role in both subject agreement and in object agreement systems.

Chomsky proposes a unified and restricted theory of (abstract) Structural Case-assignment. There is no Exceptional Case-marking nor is the sister to V (the direct object) a Case-marked position. Rather, Structural Case is assigned in only one configuration: Specifier of Agr. However, since the Case borne by the category constituting Spec of Agr does appear to be determined by the featural (Case-assigning) properties of V (for the object) and T (for the subject) and not by Agr itself, it is proposed that T˚ raises (adjoins) to AgrS˚, and V raises (adjoins) to AgrO˚. These head-movements create complex (molecular) heads which contain the φ-features of Agr as well as the Case-features of V/T adjoined to Agr. An NP (or DP) Specifier of such a complex head (of $[_{Agr}$ T + AgrS] or of $[_{Agr}$ V + AgrO]) bears the Case and agreement features of the complex head. As a result, Case and agreement are also unified–each a manifestation of the Spec-head relation, which along with the head-head relation (V-to-AgrO raising and T-to-AgrS raising) constitute the central X′-theoretic relations involved in inflectional morphology. Notice that the subject and object inflectional systems are also unified–the subject is Spec-Agr(S); (agreement is determined by the φ-features of Agr and Case by the Case-features of Tense adjoined to Agr) and similarly the object is Spec-Agr(O) (with agreement (similarly) determined by AgrO and Case by the V which adjoins to it). Properties of the CP projection remain largely unchanged from previous work (Chomsky 1986a, b)–Spec-CP is the canonical landing site of Wh-movement and C˚ (the head of CP) is construed as a mood-indicator, its features (e.g., +wh) determining the force of the AgrSP whose AgrS˚-head is selected by C˚.

Importantly, the basic clause structure raises fundamental questions concerning the proper typology of positions. In Chomsky (1981), NPs occupied two types of positions: A (argument) or A′ (non-argument), defined in relation with (potential) theta-role assignment (Chomsky 1981:34ff; see below, section 2.4). In Chomsky (1993), Chomsky seeks to eliminate the A vs. A′ dichotomy and attempts to replace it with a positional taxonomy directly related to the morphological features of lexical items (consistent with the central role played by lexical items and their morphological features within the minimalist framework).

As Chomsky notes, Agr and Tense incorporate (or represent) morphological features of the verb. These features are called "V-features" and their derivational function is to check the Agreement and Tense features of the verb that raises (adjoins) to Agr and Tense (as discussed above) and to check Case and Agreement properties of the NP/DP that raises to their Spec position. Thus, they are mediators that ensure that DP and V are properly related. More generally, V-features, being features of a lexical (L) item are called L-features. Positions, in the minimalist framework, are defined in terms of their relation to L-features. Thus, a position is L-related if it is in a local (X′-theoretic) relation to an L-feature, i.e. if it is the Spec of or the Complement of a head (X˚) bearing an L-feature.

In Chomsky (1993), a fundamental distinction between complement positions and specifier positions is assumed. Complements and specifiers belong to different domains, which are defined on the structures built by Merge and Move. The specifier of an X˚ is part of the checking domain of the X˚ wherein Case and Agreement relations between a head and a DP are established. The complement of an X˚ is part of the internal domain of an X˚ wherein theta-relations between an X˚ and an argumental DP are established.

The definitions of checking domain and internal domain of a head H can be paraphrased as follows:[14]

(12) a. The domain of H is the set of nodes dominated by the maximal projection of H, with the exception of H itself and the projections of H.
b. The complement domain of H is the part of the domain of H that is (dominated by) the sister of H; the remainder of the domain of H is the residual domain of H.

(13) a. The minimal domain of H is a subset of the domain of H, namely those nodes in the domain of H that are not dominated by another node in the domain of H.
b. The internal domain of H is the minimal complement domain of H (i.e., the sister of H); the checking domain of H is the minimal residual domain of H (i.e., the specifier of H and the adjuncts of (a projection of) H).

According to (13b), the checking domain of a head H is assumed to contain not only the Spec of H (and a head adjoined to H) but also a category adjoined to H^{max}. Spec is called a "narrowly L-related" position, while a position adjoined to H^{max} is dubbed "broadly L-related". As Chomsky (1993: 28-9) notes:

> A structural position that is narrowly L-related has the basic properties of A-positions; one that is not L-related has the basic properties of A′-positions; in particular [Spec-C], not L-related if C does not contain a V-feature. The status of broadly L-related (adjoined) positions has been debated, particularly in the theory of scrambling.[footnote deleted] For our limited purposes, we may leave the matter open.

Thus, Chomsky intends to replace the A/A′ positional taxonomy with the L-based system recognizing at least the positional types L-related, including broadly and narrowly L-related, and non-L-related.

1.6 Universality, parametric variation, and morphological strength

According to the Minimalist Program, the derivation of a sentence yields two interface representations, a PF and an LF representation, each of which is subject to the principle of Full Interpretation, requiring that an interface representation must consist of legitimate objects only. For PF, the empirical question "What are the legitimate PF objects?" will be answered by an empirically adequate theory of Universal Phonetics. At LF, each legitimate object is assumed to be a chain, including X°-chains, argument-chains, modifier chains and operator-variable chains.

If an interface representation consists of only legitimate objects the derivation yielding it is said to converge at this level of representation, e.g. "The derivation converges at PF". If a derivation D converges at both PF and at LF, then D is said to converge – period. Failure to converge is called "crashing". A derivation D can crash at PF, crash at LF, or crash at PF and at LF. Importantly, if D converges it is not necessarily the case that each interface representation is interpretable/sensical. Thus, e.g. the well known cases of nonsense poetry

[14] See Chomsky (1993: 10-13) for complete definitions of checking domain and internal domain of a head and of a head movement chain.

and sentences like *Colorless green ideas sleep furiously* consists entirely of legitimate objects but are not interpretable or are nonsensical.

The other components of the cognitive system that the syntactic component is embedded in, are performance systems, concerning (roughly) speech and interpretation. As noted, there are assumed to be two discrete performance systems: articulatory-perceptual and conceptual-intentional (Chomsky 1993:2). In accordance with this, there are two interface representations (PF and LF), each hypothesized to provide instructions to one of the two performance systems: PF to A-P, LF to C-I.

On the assumption that the human conceptual-intentional performance system is an invariant property of the species, the interface representation called "LF" must be essentially identical in all languages. By contrast, the interface representation called PF varies from language to language, as can easily be observed. The point to be made here is actually more subtle. What differs (in a restricted way) cross-linguistically among PF representations is the phoneme order, phoneme inventory, and the language-particular phonetic rules–all presented (in the data) and thereby representing no learnability problem (for humans!). The way the corresponding instructions are interpreted by the articulatory-perceptual performance system is presumably just as universal as the way the LF instructions are interpreted by the conceptual-intentional performance system.

In the Minimalist Program, it is assumed that the LF interface level is the final stage of a derivation, and that the PF interface level is the reflection of an intermediate stage in the derivation to LF. That is, at a certain point in the derivation, instructions to the articulatory-perceptual system are issued by applying a rule/operation called Spell-Out. That part of the derivation preceding Spell-Out is called overt syntax, the remainder of the derivation, i.e. that which follows the application of Spell-Out is called covert syntax.

The problem of comparative linguistics is to determine the (highly limited) formal respects in which languages can differ, and then to ascertain why language varies in the ways it does. Recall that the principle of Procrastination dictates that movements take place as late in the derivation as possible. This principle, then, has to be violated to some extent in the grammar of certain, perhaps all, languages. The question is why? That is, why isn't "as late as possible" always synonymous with "in the covert syntax"? Equivalently, given Procrastination, why isn't movement always postponed until after Spell-Out, in which case not a single language would have overt movement? In other words, why doesn't the A-P system spell out what used to be called D-structure in Government and Binding terms?

The only possible answer to this question within the highly restricted Minimalist framework is that overt movement applies, i.e. Procrastination must be violated , in order to ensure convergence at the PF interface level. In other words, certain objects that would be illegitimate at PF must be eliminated in overt syntax. Adhering to the minimalist assumptions made above, it may be the case that certain inflectional features count as illegitimate objects at PF. These features, then, have to be checked and eliminated in overt syntax, through a process of movement of heads and phrases to positions in the functional domain.

Notice that this view is quite consonant with earlier analyses. For example, *I wonder* [COMP (+wh)] *John likes what* is the wrong way to say – i.e. the wrong sound/PF in English. Rather, a COMP selected by *wonder* requires a local wh-phrase–where locality can be obtained by movement, viz. *I wonder what John likes*. Thus, in English a [+wh]COMP without wh is an illegitimate PF object.

The surprising aspect of this mechanism is that the class of illegitimate objects at PF is not universalized but rather varies cross-linguistically. If there were universality, overt syntax

would be largely, perhaps completely identical in all languages. As we know, there are very distinct differences in word order between even closely related languages such as English and French (Pollock 1989).

This, then, appears to be the locus of syntactic, i.e., word order parametrization between languages: an inflectional feature may or may not be visible as an illegitimate object at PF. In English, a +wh COMP without a wh-specifier in its checking domain is an illegitimate PF object. Thus, if in a derivation D we were to Spell-Out

(14) *I wonder* [$_{CP\ +wh}$ C° [$_{IP}$ *John likes what*]]

The derivation crashes at PF since C° lacking a wh-specifier is an illegitimate PF object. Categories that are visible as illegitimate objects at PF will have to be eliminated in overt syntax. By the principle of Procrastination those that are not visible at PF will not be eliminated in overt syntax, i.e., "if you don't have to move for convergence, then movement is prohibited". Features that are visible (thus: potentially illegitimate) at PF are called strong*;* features that are invisible (thus: inoffensive) at PF are called weak.[15]

A minimal assumption is that the strong/weak distinction constitutes the only parametric variation among languages. This implies that parametric variation is restricted to certain features (e.g., +wh) of functional categories (e.g. C°) (Fukui and Speas 1986). Furthermore it implies that there are no directionality parameters, such as directionality of government.[16] The latter implication is supported empirically by Kayne (1994), who argues that movement is invariably leftward.

1.7 Directionality and word order

The structure building process of iterative (cyclic) generalized transformational application nowhere specifies the linear order of head, complement, specifier, and adjunct. Superficial cross-linguistic examination suggests that languages may differ with respect to the linear order of these elements. In the tradition of generative grammar, the attested variation is described in terms of a parametric option: heads may govern to the left or to the right. A head that governs to the left takes its complement to the left in the initial representation, yielding an OV order at D-structure.

In the minimalist approach, a directionality parameter is no longer available. First, parametric variation must be expressed in terms of the features of functional heads only. A directionality parameter would therefore not suffice to account for the ordering of elements in the lexical domain. Second, government no longer plays a role in the minimalist approach. Therefore, it is unclear whether a directionality parameter could be reduced to properties of such an independently defined grammatical relation. Third, a directionality parameter would be redundant, since much of the word order variation can be accounted for by the interaction of overt and covert movement.

[15] See Koster 1986 for earlier use of this terminology in connection with word order variation.

[16] In fact, government has no formal status in the Minimalist Program. For example, Case assignment is reduced to feature checking in a specifier-head configuration (see also section 2.6 below). The consequences for the Empty Category Principle, which incorporates the notion proper government*,* have not been fully explored.

Kayne (1994) presents empirical evidence indicating that movement into the functional domain is invariably leftward. The evidence consists in what we do *not* find, in comparing cross-linguistic movement phenomena. Thus, we can hypothesize from the apparent absence of Wh-movement to the right that the specifier position of CP is always to the left. Similarly, there appear to be no known cases where verb movement changes a verb-complement order from VO to OV, which suggests that verb movement to the right does not exist. Hence, the functional projections hosting V-features are all assumed to be head initial. Also, the subject precedes the object in almost all languages of the world (Greenberg 1963, Universal 1). Assuming, in connection with this, that AgrSP is hierarchically higher than AgrOP, it also follows that the specifier of AgrSP is situated to the left. Likewise, if the complement of a preposition is extracted, the complement always appears to the left of the preposition, never to its right. Again, this suggests that licensing positions, i.e. specifier positions, under our assumptions, are on the left hand side.

These empirical observations are presented more fully in Kayne (1994), who additionally argues that lexical projections are universally head initial as well. This is an attractive hypothesis, considering the empirical evidence for the universal structure of the projections of the functional domain. However, empirical evidence in support of this hypothesis is considerably more difficult to obtain, in view of the fact that the observable word order reflects an intermediate stage of the derivation, i.e., one never knows whether the constituents are in a basic position or not.

Kayne (1994) also presents conceptual argumentation in support of the idea that all phrases are head initial, proposing that asymmetric c-command invariably maps into linear precedence. In order for this mapping to be successful, it must be possible to express the relations between the nodes of a phrase marker that asymmetrically c-command into a set of ordered pairs <x,y> of the terminal (lexical) elements dominated by these nodes. The pairing of two terminal elements x,y thus expresses a relation between x and y. Kayne proposes that the set of ordered pairs of these relations must express a linear ordering, i.e. a total, transitive, and antisymmetric ordering.

Thus, according to this proposal it must be possible to read the relation of each terminal element to all other terminal elements off of the set of ordered pairs. Crucially, these relations must be antisymmetric, i.e., it is excluded that two terminal elements L each other, where L stands for the relevant relation between these two elements.

The axiom that the set of ordered pairs of terminal elements (derived from the set of relations between the nodes of a phrase marker that are in a relation of asymmetric c-command) is a linear ordering of the terminal elements is called the Linear Correspondence Axiom (LCA).

In addition to the LCA, Kayne proposes that the relation expressed by the pairing of terminal elements is a precedence relation. We will refer to this hypothesis as the Extended Linear Correspondence Axiom (ELCA).[17]

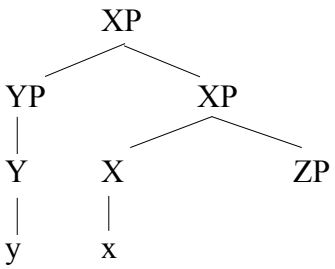
[17] Kayne (1994: section 5.3) derives the definition of the relation between the terminal elements in a phrase marker as a precedence relation from the hypothesis that every phrase marker contains a root node dominating all other nodes except itself. On the assumption that this root node also dominates an abstract terminal element *a*, which, as Kayne argues, has to precede all other terminal elements of the phrase marker, it follows that the linear relation between *a* and the other terminal elements is also a precedence relation. Hence, the linear relation between terminal elements must always be a precedence relation. For empirical justification of the Extended LCA, see above.

Kayne shows that the adoption of the LCA explains many basic facts of phrase structure, such as binary branching and endocentricity. In this respect, the LCA is compatible with the mechanism of generalized transformations as presented in section 1.2. It follows from the ELCA that adjunction always takes place on the left hand side.

In some respects, however, the LCA appears to be too restrictive, as Kayne notes. In fact, the LCA excludes adjunction of specifiers and adjuncts. Kayne therefore modifies the definitions entering into the notion c-command in order to allow adjunction of specifiers. He argues, however, that adjunction of adjuncts (i.e. in addition to adjunction of a specifier) is excluded.

To see why adjunction of specifiers is difficult, consider the tree structure in (15), where y and x represent terminal elements:

(15)



Assume the following definition of c-command:

(16) α c-commands β iff every γ that dominates α dominates β.

C-command is asymmetric where, for α c-commanding β, β does not c-command α.

In (15), YP asymmetrically c-commands X and XP asymmetrically c-commands Y. YP dominates the terminal element y, and X dominates the terminal element x. The relation between YP and X therefore can be expressed in the ordered pair of terminal elements <y,x>. But since XP dominates x and Y dominates y, the ordered pair of these terminal elements <x,y> is also part of the set of ordered pairs expressing the relations between YP, XP, Y, and X. So now this set contains <y,x> and <x,y>. Hence, the relation between x and y (i.e., between a head and its specifier) is not linear, because it is not antisymmetric.

To solve this problem, the pair <x,y> or the pair <y,x> must be excluded. This can be achieved if either YP does not c-command X (eliminating the pair <y,x>) or XP does not c-command Y (eliminating <x,y>). Kayne proposes to modify the definition of c-command in such a way that XP no longer c-commands Y. This can be done by excluding segments from the definition of c-command, assuming the higher XP and the lower XP in (15) to be two segments of the same category:[18]

(17) α c-commands β iff (i) α and β are not segments, and
(ii) α excludes β, and
(iii) every γ dominating α dominates β.

(18) α excludes β if no segment of α dominates β.

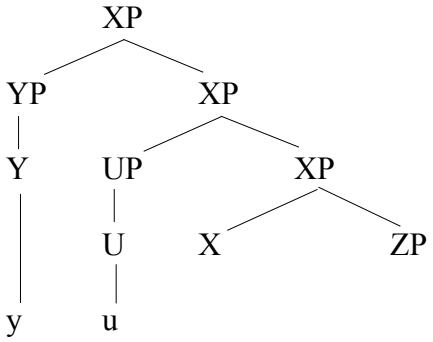
[18] Following May 1985:57, Chomsky 1986b:7. For critical discussion of this segmental theory of adjunction see also Epstein 1987, 1989, 1991, 1993, to appear, Lasnik and Saito 1992, Williams 1988, and Poole, forthcoming.

In (15), XP is a segment, hence does not c-command Y by clause (i) of the definition of c-command in (17). This gives the desired result that the relation between *x* and *y* in (15) is described by <y,x>, hence is a linear relation (hence, following Kayne, a precedence relation).

Notice that the fact that the higher XP is a segment of the lower XP suffices to exclude that the lower XP c-commands Y. Since the higher XP is a segment, the lower XP does not exclude Y, and the c-command relation is barred by clause (ii) of the definition of c-command in (17) (cf. Kayne 1994, note 2 in chapter 3).

To see why (multiple) adjunction is difficult, consider (19):

(19)



In (19), not only the specifier UP, but also the adjunct YP is adjoined to XP. YP asymmetrically c-commands U, yielding <y,u>, and UP asymmetrically c-commands Y, yielding <u,y>. Hence the relation between the terminal elements y and u is not antisymmetric and therefore (19) is not allowed by the LCA.

Kayne concludes that multiple adjunction (i.e., adjunction of an element in addition to adjunction of a specifier, as in (19)) is universally impossible. It follows that adjunct elements, such as adverbs, can be present only as specifiers. Thus, for every adjunct there must be a head in the structure creating the required specifier position.

It is important to note that the ELCA does not necessarily follow from the LCA. The ELCA merely interprets linearity in the mathematical sense in one-dimensional terms, yielding precedence. Chomsky (1995a) objects to the way Kayne (1994) employs virtual categories (such as intermediate bar nodes) to determine the relation between terminal elements. He therefore rejects the LCA as proposed by Kayne, while accepting the word order generalizations of Kayne in full.

2. Introduction to the studies in this volume

The studies collected in this volume cover a range of issues that have taken center stage in the Minimalist Program: (i) the basic operations that create phrase structure (Merge and Move), (ii) licensing of grammatical categories (movement and feature checking), (iii) economy conditions and restrictions on movement, (iv) the nature of phrase structure and the inventory of functional categories, (v) inflectional morphology and the nature of parametric variation, and (vi) the relation between hierarchical structure and linearization. In the remainder of this introduction, we will present these issues and discuss the contributions that the various papers make in the respective domains.

2.1 Building up phrase structure

As we have seen in section 1.2, Chomsky (1993) departs from previous analyses, in that the rules or mechanisms deriving phrase structures are considered to operate in a bottom-up fashion. In the tradition of generative grammar, phrase structure is described by a system of rewrite rules, which take a phrase and analyze it into its constituent parts:

(20) S → NP VP
VP → V NP

The rewrite rules can be regarded as the generative counterpart of the immediate constituent analysis of Bloomfield (1933), and of the traditional analysis of a sentence into its parts of speech.

Within generative grammar, there has been a development of category specific rewrite rules, such as the ones in (20), to the category-neutral rewrite rules of the X-bar Theory, illustrated in (21) (see Chomsky 1970):

(21) XP → YP X'
X' → X° ZP

However, this development did not change the top-down character of the phrase structure rule system.

The X-bar Theory holds that phrases are structured according to the category-neutral rewrite rules of the type in (21). The rewrite rules convey two types of information. They express that there is a projection of categorial features running from the head of a phrase to the maximal projection of the phrase (the projection line X°-X'-XP). Second, they express that the nodes making up the projection line are of different levels (X° being the head, X' the intermediate projection, and XP the maximal projection).

Thus, X-bar Theory basically asks (and anwers) the question: If we have a phrase, what does it consist of? Chomsky (1993) turns the tables, and asks: If we have a syntactic object (a head or a phrase), what do we get when we integrate that object in a syntactic structure?

For example, if we want to integrate a head, say V, into a syntactic structure, we need to combine it with something, its complement, say DP, and the combination of V and DP yields a new syntactic object, V'. Hence the mechanisms Merge and Move describe how two syntactic objects are combined, and what syntactic object the combination yields.

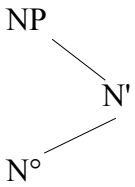
In Chomsky (1993), X-bar Theory is not eliminated. The operations Merge and Move are subject to the principles of X-bar Theory:

> The computational system takes representations of a given form and modifies them. Accordingly, UG must provide means to present an array of items from the lexicon in a form accessible to the computational system. We may take this form to be some version of X-bar theory. The concepts of X-bar theory are therefore fundamental. In a minimalist theory, the crucial properties and relations will be stated in the simple and elementary terms of X-bar theory. (Chomsky 1993: 6)

This suggests that the top-down rewrite system and the bottom-up Move/Merge system are just notational variants. However, certain discrepancies do exist. For example, a descriptive

system that uses the top-down rewrite system of X-bar theory must introduce the intermediate X'-level, even if no specifier or complement exists. Thus, NP is always rewritten in such a way that the structure in (22) results:

(22)



In the bottom-up system, N' can only be the result of combining N° with a complement. In fact, if no complements and specifiers exist, it may be impossible to tell N° and NP apart.

What this example illustrates is that in the bottom-up system, phrase level is a relational concept, concomitant with movement processes, rather than an absolute concept given by rules. Therefore, the shift to a bottom-up system in Chomsky (1993) is more radical than the above quote suggests. It is, however, a perfectly natural shift, given the tendency in generative grammar since the 1970s towards eliminating rules altogether.[19]

The question then remains what Chomsky means by "the concepts of X-bar theory" which are taken to be "fundamental". This must relate to the two types of information that the rules of X-bar theory convey: categorial status and phrase structure level status. We want to maintain that there is a notion of a head that projects, and that the categorial features of the head are also present on the projections of the head. Secondly, we need to have an algorithm that tells us about the phrase structure level of the syntactic object resulting from the operation Merge/Move.

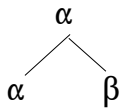
The issue of projection is discussed in much detail in Chomsky (1995a,b). Chomsky argues that if α adjoins to β, only β projects its features (1995b:243f, 256f). In other words, γ, the combination of α and β, can be defined as the set of nodes containing α and β, and a bundle of features present on γ, which derive from β, and which are represented as the label of γ:

(23) $\gamma = \{\text{LABEL}\{\alpha,\beta\}\}$

We will pass over the details of the argumentation here, as they are not relevant to the articles in this volume.

Note incidentally that the question of which node projects is relevant only to the operation Move, as the operation Merge is defined in such a way that what projects counts as the head (Chomsky 1995b:244). In other words, Merge combines a head and a nonhead:

(24) Merge



$\alpha = X°$

$\beta = XP$

[19] In Chomsky 1995b:256, the relation of the present theory with the traditional X-bar theory is made explicit. X-bar theory, it is said there, consists of "stipulated properties" which we can no longer make recourse to, "so we hope to show that the conventional assumptions are in fact derivable on principled grounds".

With the operation Move, a typical case is where two nonheads are combined (movement to specifier position or adjunction of XP), or (in the case of head movement) two heads. Hence the relation is symmetric, and the question which element projects its features becomes relevant.

The second type of information conveyed by the rewrite rules of X-bar theory concerns the phrase structure level of the various subtrees in the structure. In other words, how can we derive that the higher α in (24) is X' or XP, rather than X°?

Here Chomsky (1995b:242) resorts to a relational definition of phrase structure, introduced by Muysken (1982). Muysken proposed that phrase structure level is not an inherent property of subtrees, but a relational property, to be expressed in terms of the features [maximal] and [projection]. These two features yield the following combinatorial possibilities:

(25) a. [+maximal,+projection] XP
b. [+maximal,-projection] [?]
c. [-maximal,+projection] X'
d. [-maximal,-projection] X°

As can be seen, these combinatorial possibilities yield the three conventional X-bar levels. ((25b) represents the case of a head that does not project. Possibly clitics are a case in point.) Importantly, however, the node immediately dominating a head is not necessarily X' in this system. X° projects an X' only if X' is not the maximal projection of X° (i.e., if X' is immediately dominated by another projection of X°). If the immediate projection of X° is the highest node in the projection line of X°, its features are [+maximal,+projection], corresponding to XP.

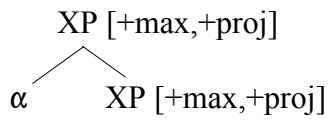
In this system, unlike in the traditional X-bar theory, X°, X', and XP are derived notions. In the phrase structure representation proposed in Chomsky (1995a,b), these notions do not occur (see (23)). Notice, however, that the relational definition of phrase structure level based on (25) involves two features only because we want to derive three levels of phrase structure. It has been proposed several times that there is no syntactic distinction between XP and X' (Stuurman 1985, Hellan 1991, E. Hoekstra 1991, Zwart 1992b, Kayne 1994). If this is correct, we can replace (25) by (26):

(26) a. [-projection] X°
b. [+projection] XP

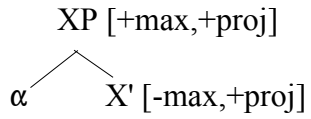
In (26), we seem to have reached a fundamental property of phrase structure: there is a distinction between heads and phrases. It is not clear that (25) reaches the same fundamental level. If not, the system in (25) is just one way of capturing a traditional, essentially stipulative property of X-bar theory, namely that we must distinguish three levels of phrase structure.

The question of the existence of an intermediate X' level is relevant for at least two issues. First, Chomsky (1995b) maintains that there is a fundamental distinction between substitution and adjunction. Both substitution and adjunction are subcases of the operation Move. Adjunction differs from substitution in that adjunction of α to β yields a syntactic object of the same phrase structure level as β, whereas substitution yields a syntactic object that is one phrase level higher than β:

(27) a. adjunction



b. substitution



This distinction between substitution and adjunction cannot be expressed if phrase structure level is defined in terms of [±projection] only.

Notice that in (27a) the characterization of the lower XP as [+maximal] is not straightforward (as there is a second XP immediately dominating it). Therefore, the two XPs in (27a) are characterized as two "segments" of a single maximal projection (as in May 1985, Chomsky 1986a).

The choice between (25) and (26) is also relevant if we wish to construe an algorithm describing the features of γ, resulting from the combination of α and β (via Merge or Move). As argued in Zwart (1992b, 1993:25), such an algorithm would be more elegant if we start from the feature system in (26) (i.e., if we do not distinguish XP and X'). Informally speaking, Zwart proposes that γ takes its categorial features from the head (say α), and its phrase structure level from the nonhead (β):

(28) If β^n is combined with α^m, the result is α^n

(where *m,n* specify phrase structure level, and α,β categorial features)

The algorithm in (28) covers the following cases (where Y = α):

(29) a. XP + Y° = YP
b. X° + Y° = Y°
c. XP + YP = YP
d. X° + YP = Y°

Of these, only (29d) does not exist. This, however, follows from the definition of Merge. (29d) involves a head, which can only be combined with a nonhead by projecting (see (24); we could stipulate that *m* in (29) must not be higher than *n*). The other cases are well described by the algorithm in (29), assuming no distinction between maximal projections and intermediate projections.

Kayne (1994) presents a view of phrase structure that differs significantly from that of Chomsky (1993, 1995a,b). Kayne takes the difference between heads and nonheads to be fundamental, and therefore rejects the distinction between immediate projections (X') and maximal projections (XP). Like Chomsky, Kayne wants to derive the well-known properties of X-bar theory from more fundamental notions. However, unlike Chomsky, Kayne wants to relate these notions to the linear order of the terminal nodes (see 1.7).

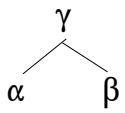
We will return to this issue in section 2.6. Here we will briefly indicate where Chomsky's and Kayne's views on phrase structure diverge. First, Kayne's system requires vacuous projection, which is not allowed in Chomsky's approach. Second, Kayne's system prohibits multiple adjunction, which is not prohibited in the minimalist approach.[20]

As we have seen, Kayne proposes to derive the linear relation between two elements α and β from the hierarchical relation between α and β (the LCA, see section 1.7 above). This relation must be asymmetric, and is defined in terms of c-command (cf. Kayne 1994: 16):

(30) α c-commands β iff (i) α and β are not segments, and
(ii) α excludes β, and
(iii) every γ dominating α dominates β

In other words, the relation between α and β is well-defined only if there is a hierarchical relation between α and β (or between the projections of α and β) such that α c-commands β and β does not c-command α. Consider now the case where α and β are sisters:

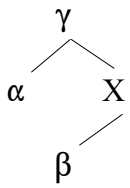
(31)



Here, α c-commands β and β c-commands α. Consequently, there is no hierarchical asymmetry between α and β, and the relation between α and β is not well-defined. Structures like (31) are excluded in Kayne's system.

There are two ways to establish a hierarchical ordering of α and β in (31). The first is to embed β deeper:

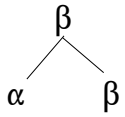
(32)



In (32), α c-commands β, but β does not c-command α. Notice that the presence of X is only motivated by the requirement that α and β be ordered. X does not result from the combination of β with a complement, as in Chomsky's system. Consequently, Chomsky prohibits structures like (32), whereas Kayne requires them.

A second way of fixing (31) would be to consider γ a segment of β:

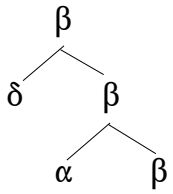
(33)



[20] Chomsky 1995b:323 argues that adjunction is a severely restricted option in the minimalist approach, but multiple adjunction of specifiers is not. Recall that in Kayne's system, a specifier is technically an adjunct (a sister of XP), so that multiple specifiers in Chomsky's system are multiple adjuncts in Kayne's system.

Being a segment, β does not c-command α by (30i). Notice that adding another segment would yield new problems, as discussed in section 1.7:

(34)



Here the problem is that δ and α c-command each other, as neither is dominated by β. As we have seen in 1.7, Kayne adopts Chomsky's (1986b:7) modification of the definition of dominate here:

(35) α dominates β only if every segment of α dominates β

Thus, Kayne (1994) disallows multiple adjunction structures of the type in (34), whereas Chomsky (1993, 1995b) permits them.

The phrase structures allowed in Kayne's system are highly regular and maximally simple. Whenever a phrase seems to have two adjuncts adjoined to it (for example, a specifier and an additional adjunct, as in (34)), the second adjunct is assumed to be in the specifier of a higher functional projection. This higher functional projection is present only because the theory of possible phrase structure forces it to be there, not because its head hosts some inflectional feature. Again, this is a considerable departure from traditional thinking (within generative grammar) about phrase structure (see also the discussion of functional projections in sections 2.4-5 below).

The question of permissible phrase structures is taken up in this volume in VANDEN WYNGAERD's article. VANDEN WYNGAERD applies Kayne's system to the morphosyntax of Dutch past participles, which consist of a prefix (*ge-*), a verb stem, and an inflectional ending. Treating these three elements as being organized in a phrase structure that meets the requirements set by Kayne (1994), VANDEN WYNGAERD accounts for several features of past participle constructions in Dutch. Among these is the phenomenon that participles that have an infinitive in their complement domain are turned into infinitives in various Continental West Germanic dialects (the Infinitivus Pro Participio effect). VANDEN WYNGAERD also addresses the question of nonbranching projections. If these do not exist, as Chomsky (1995a:399) holds, unergative verbs can be distinguished from unaccusative verbs only if they are in fact transitive (as argued independently by Hale and Keyser 1993). VANDEN WYNGAERD's analysis of Dutch past participles presents support for this analysis of unergatives as transitives.

The processes of Merge and Move raise a number of other questions which are discussed in the articles in this volume.

The operations Merge and Move are technically identical: in both cases, a phrase marker α is adjoined to another phrase marker β. In both cases, β has to be the root (the top node) of its phrase marker (the extension condition or strict cycle condition; see 1.1.3 and 2.3). The difference between Merge and Move is that in the case of Move, α is already contained in β, whereas in the case of Merge, α and β are two independent phrase markers. We may think of Merge as taking elements from the Lexicon as input (or a selection of elements from the Lexicon, which Chomsky 1995b:225 calls the numeration). Chomsky (1993:22) assumes that

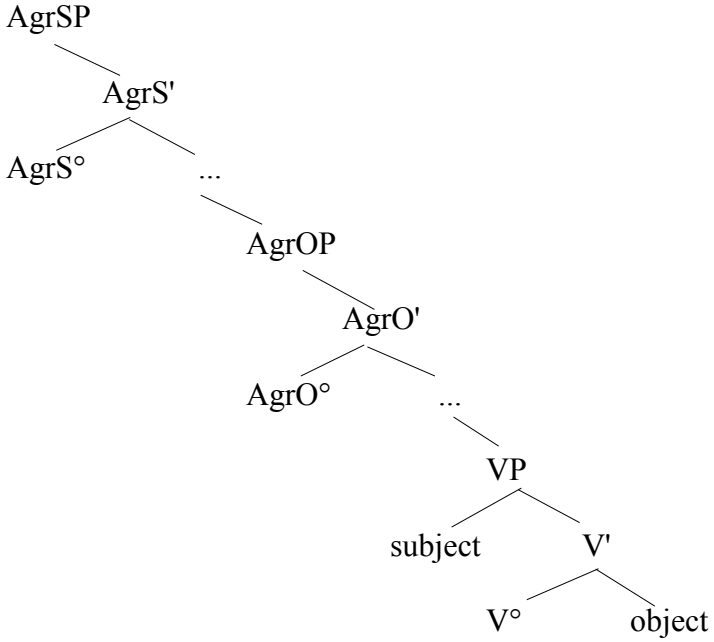
no elements may be taken from the Lexicon after Spell-Out. Thus, Merge is restricted to apply in overt syntax only. This asymmetry between Merge and Move is addressed in the article by GROAT AND O'NEIL.

GROAT AND O'NEIL seek to eliminate the asymmetry by dissociating movement from Spell-Out. A standard assumption in generative grammar is that an element α, when it moves, is spelled out in the position where α ends up as a result of the movement (i.e. in its ultimate landing site). If we think of movement as creating a chain consisting of a chain-head (the landing site of the movement) and one or more traces, the standard assumption amounts to saying that the head of a chain is the element that is spelled out. Chomsky (1993:34-35) takes a trace of the moved element α to be a copy of α. GROAT AND O'NEIL now propose that when α moves, either α or the copy of α (the trace) can be spelled out. In this proposal, languages do not differ with respect to the timing of movement, but with respect to the chain-member (copy) that is spelled out (pronounced). This makes it possible to restrict the operation Move to overt syntax, just like the operation Merge (see also section 2.7 below).

Another question raised in connection with the structure building operations Merge and Move concerns the structure of the functional domain. In traditional X-bar theory, the structure of the functional domain is encoded in the rewrite rules. Thus, since C selects IP, IP must be contained in CP. If Merge operates freely, the ideal case in a minimalist approach, generalizations about the dominance relations among the functional projections must be derived in some way.

One such generalization is that the functional projection licensing the object (AgrOP) is contained in the functional projection licensing the subject (AgrSP):

(36)



Chomsky (1993:18) derives this generalization by imposing a locality condition on the movement operations that take the subject and object from their VP-internal positions to the specifier positions of AgrSP and AgrOP, respectively. These conditions are discussed in section 2.3 below. The question whether these conditions exist, and, if so, how they should be formulated, is addressed in the articles by POOLE and ZWART.

Baker (1988) has proposed that the hierarchical relations within the functional domain mirror the morphological structure of inflected forms. Thus, TP is supposed to be contained

in AgrSP when the tense morpheme appears closer to the verb stem than does the agreement morpheme. This Mirror Principle is discussed further in section 2.4 below, and plays an important role in the papers by THRÁINSSON and SOLÀ. Discussion of the question of which functional projections must be assumed to exist in the grammar of a given language will also be postponed until section 2.4 below.

A final question regarding the operation Merge concerns the position in which the subject is generated (or, in minimalist terms, the moment in the derivation at which the subject is merged with the existing structure). Chomsky (1993:8) adopts the by now standard assumption that the subject is generated in the specifier position of VP (merged with V') (see (36)). In Chomsky (1995b:331), this assumption is essentially maintained (albeit that VP is now assumed to contain two VP projections, the subject being generated in the specifier position of the higher VP).[21]

The assumption that the subject is generated in the specifier of the VP is not challenged in the articles in this volume. We mention the issue, however, because this asumption is of crucial importance to several articles, including KITAHARA's (in which the trace of the subject in Spec-VP plays a role in the description of scope asymmetries), and all those articles that address the locality conditions on movement of subjects and objects (FERGUSON, JONAS, POOLE, and ZWART).

Next to the operations Merge and Move, Chomsky (1993:15) appears to propose a third type of operation forming syntactic objects, Form Chain. Form Chain plays a role in successive cyclic movement constructions, such as (37):

(37) *Who$_i$ did you say* [$_{CP}$ *t$_i$ that Bill had kissed t$_i$*] ?

In (37), the syntactic object created is the chain (*who,t,t*).

Successive cyclic movement typically involves movement out of a local domain (CP in (37)) via intermediate landing sites (Spec-CP in (37)) (Chomsky 1973, 1986b). In the minimalist framework, both the movement to the intermediate landing site and the movement to the final landing site must be triggered by feature checking requirements.

Here, a potential problem arises. It is reasonable to propose that movement to both landing sites is triggered by a requirement to check the wh-features of the wh-element (*who* in (37)). For the movement to the matrix CP, this is unproblematic. We may assume that the matrix C has [+wh]-features, since the clause as a whole has interrogative properties. So we may assume that the wh-feature of *who* is checked in the Spec-CP of the matrix clause. But there is reason to believe that the wh-feature of *who* is also checked in the Spec-CP of the embedded clause. In many languages the embedded complementizer is affected in some way by the long distance movement, either by special morphology or because a different complementizer is used in that situation. In addition, there are languages in which the Wh-movement can be partial, i.e., it moves the wh-element to the embedded CP only (in that case, an interrogative scope marker appears in the matrix Spec-CP). Both phenomena suggest that checking of the wh-features in the embedded CP cannot be excluded.

The problem, then, is the following. Suppose movement of *who* in (37) to the embedded Spec-CP is triggered by the requirement that the wh-feature of *who* be checked. Then, *who*'s wh-feature will be checked in the embedded clause, and (assuming that checing invariably

[21] For the VP-internal subject hypothesis, see, among others, Kitagawa 1986, Fukui and Speas 1986, Koopman and Sportiche 1991, Burton and Grimshaw 1992, McNally 1992, and references cited there.

entails deletion) there will be no feature left to check in the Spec-CP of the matrix clause. Thus, checking the feature in the embedded Spec-CP will remove the trigger for further movement to the matrix Spec-CP. Suppose on the other hand that movement of *who* to the embedded Spec-CP does not result in checking of the wh-features of *who*. This will make it possible for *who* to move on to the matrix Spec-CP, assuming that its features will be checked there. However, now it is unclear what the trigger is for moving *who* to the embedded Spec-CP in the first place, since the movement does not result in feature checking (and should be disallowed by Greed).

Underlying this problem is a conflict between the two economy requirements that the operation Move is subject to. These requirements are that steps in the movement process should be as few and as short as possible. The requirements are conflicting, because the number of steps decreases with the distance of the steps. In (37), *who* could have moved to the matrix Spec-CP in one step, satisfying the fewest steps requirement, but violating the shortest steps requirement.

To solve these problems, Chomsky (1993:15) introduces the operation Form Chain, which creates (37) in a single step, forming the chain (*who,t',t*) in the process. The nature of the operation is admittedly unclear. We will discuss it further below, in section 2.3. In this volume, POOLE and ZWART discuss Form Chain extensively.[22]

2.2 Licensing: movement and feature checking

A basic theoretical and descriptive problem in syntax is to account for the fact that only certain syntactic positions are possible argument positions and different types of arguments are licensed in different positions. Thus subjects can occur in certain positions where objects cannot occur and vice versa. In the Government-Binding framework it is the task of Case theory to account for facts of this kind.The basic claim is that every overt DP (or NP) must be assigned (abstract) Case (cf. e.g. the formulation of the Case Filter in Chomsky 1981: 49 and elsewhere, which assumes the basic idea of (abstract) Case usually attributed to Vergnaud (see, for instance, Rouveret and Vergnaud 1980, Vergnaud 1982)). Within the GB framework, then, one consequence of this is that if an argument is base-generated in a position where it cannot be assigned Case, the derivation will be illicit and the string predicted to be ungrammatical unless the argument is moved to a Case position (or passes through such a position in the course of the derivation).

As described in section 1.3 above, phenomena of this sort are handled in a somewhat different fashion in the Minimalist Program. Under the strict lexicalist approach assumed, arguments emerge from the Lexicon with full morphological specification, including Case features. The standard minimalist account is that the Case features (like all other morphosyntactic features) need to be checked in the course of the derivation, either by PF, if they are "strong", or in LF, if they are "weak" (see section 1.6 above and section 2.5 below). Thus Case assignment is replaced by Case checking. Furthermore, while Case assignment in GB theory is typically explained by reference to some notion of government, Chomsky (1993) assumes that Case checking uniformly takes place in a Spec-Head relationship between an

[22] Chomsky 1995b abandons the Form Chain analysis of (37), and proposes that checking of the wh-features of *who* in the embedded Spec-CP does not automatically lead to elimination of these wh-features. Not being eliminated, the features provide a trigger for further movement.

overt DP (NP) and a functional head (cf. section 1.3 above). More specifically, objects will have their Case checked in a Spec-Head relationship with an AgrO head and subjects will have their Case checked in a Spec-Head relationship with the functional head AgrS (or possibly T, cf. e.g. Chomsky 1993: 7 and 45 (fn. 11)). This is a very strong claim that raises a number of interesting questions, divisible into at least two groups: Questions having to do with universal aspects of the feature checking approach and questions about cross-linguistic (parametric) variation in feature checking. Both types of questions are discussed in considerable detail in this volume. Here we will mention a couple of examples.

First, note that under the standard minimalist account just described, it is assumed that some sort of raising is involved in the Case checking of subjects and objects in all languages: Either the arguments raise overtly to the relevant Spec-AgrP position (or Spec-TP), thereby checking the (strong) Case features by PF, or else the relevant raising and feature checking (checking of weak features) will take place in LF.[23] As discussed in section 1.6 above and section 2.5 below, the basic idea is that languages will vary with respect to which movements are overt and which ones covert, but all languages are assumed to have subject and object raising at some level for the purposes of morphological checking. This means, for instance, that an English sentence without overt object shift like *John never read the book* and a corresponding Icelandic sentence with overt object shift (i.e. *Jón las bókina aldrei*, lit. 'Jón read the book never') will have the same LF-representation, a unifying and welcome result given their apparent synonymy. One question is, then, whether we can get some explanatory mileage out of this purported similarity among languages, i.e. whether this way of approaching Case can be independently motivated, i.e. used to explain something else.

This is exactly the point made in KITAHARA's paper in connection with scope interpretation of quantified constructions. KITAHARA argues that core cases of scope interpretation phenomena that have been analyzed by appeal to the LF-rule of Quantifier Raising proposed by May (1977, 1985) are accounted for by the Case checking operations independently needed in the Minimalist Program. KITAHARA assumes (with Chomsky and Lasnik 1993) that each feature checking operation forms a distinct chain and formulates the following Scope Principle:

(38) A quantifier X may take scope over quantifier Y iff X c-commands a member of each chain associated with Y at LF.

Given this, plus standard minimalist assumptions about feature checking, KITAHARA accounts for the fact that (39a) is scope-ambiguous (i.e. *someone* can have scope over *everyone* or vice versa) whereas (39b) is not (*who* must have scope over *everyone*):

(39) a *Someone saw everyone.*
b *Who saw everyone?*

No rule of Quantifier Raising (QR) is needed to account for this difference in scope interpretation, nor the other cases that KITAHARA discusses. This is obviously the kind of

[23] The exact formulation of feature checking is not the same in all of Chomsky's minimalist work, but it is not necessary to go into the details here. The reader can compare the formulations in Chomsky 1993, 1995a and 1995b. Most of the papers under discussion here take the formulations in Chomsky 1993 and/or 1995a as points of departure, although some do incorporate references to ideas first discussed in Chomsky 1995b.

result that (if empirically adequate) every theoretical linguist (or more generally, anyone interested in explanation in any domain) dreams about: An analysis that is specifically designed to account for particular phenomena turns out to provide an explanation for apparently independent phenomena elsewhere and provides a solution to a difficult puzzle. Here the account is of particular theoretical interest since it follows from a central assumption in the Minimalist Program, something which differentiates it from previous accounts, namely the theory of morphological feature checking. KITAHARA's account raises the important question of whether the widely assumed rule of QR may be completely eliminable from grammar, an attempted reduction of or minimization of operation types generally sought within the (aptly named) Minimalist Program. That possibility is raised in his paper, but he leaves the final answer to future research.

Another interesting issue raised by the generalized Spec-Head approach to Case checking is how to account for cross-linguistic differences in the licensing of subjects and objects. One aspect of this variation is manifested in the so-called "multiple subject positions" found in some languages but not others. This difference is believed to be responsible for the fact that some languages allow the so-called ``Transitive Expletive Construction'' while others do not. This is illustrated with examples from the closely related languages Icelandic and Swedish in (40):

(40)	a	*...að*	*það*	*hefur*	*einhver stúdent*	*sennilega*	*stolið*	*smjörinu.*	(Ice)
		that	there	has	some student	probably	stolen	the butter	
		'...that some student has probably stolen the butter.'							
	b	**...att*	*det*	*har*	*någon student*	*antagligen*	*stulit*	*smöret.*	(Swe)
		that	there	has	some student	probably	stolen	the butter	

If we assume the VP-Internal Subject Hypothesis, it appears that the logical subject *einhver stúdent* `some student' (the associate of the expletive) has moved out of the VP in (40a) as it precedes the sentential adverb *sennilega* `probably', which is standardly assumed to be left-adjoined to VP in constructions of this sort. Under the minimalist approach, this means that the subject must be able to check some feature in this position in Icelandic. The corresponding derivation is unacceptable in Swedish, witness the ungrammaticality of (40b). Assuming that the position of the overt expletive, namely the position immediately preceding the finite verb, is the canonical subject position in both Icelandic and Swedish, we can describe the difference between the two languages in terms of the number of subject positions licensed.[24] In that sense Icelandic has an extra subject position which Swedish lacks. The question is how to account for this difference and this is one of the main questions that JONAS discusses in her paper. To have any explanatory power, such an account should link two or more phenomena. Otherwise it would be ad hoc.

Following a line of argument also found in work by Jonas and Bobaljik (1993) and others, JONAS maintains that the extra subject position licensed in languages like Icelandic is Spec-

[24] JONAS argues against the possibility that the position of the overt expletive is a topic position rather than a subject position, mentioning arguments given by Ottósson 1989 indicating that the overt expletive is quite different in nature from typical Spec-CP (i.e. Topic) elements.

TP and the position where the overt expletive occurs is Spec-AgrSP.[25] Her main innovation is to link the licensing of Spec-TP to overt verb movement to T, claiming that a bare T does not project a specifier. She points out that only those Scandinavian languages that show general verb movement in embedded clauses (namely Icelandic and, apparently only dialectally, Faroese) are those which allow the transitive expletive construction. In addition, she maintains that there is evidence for this extra subject position in other constructions in these languages, including raising expletive constructions, but not in languages lacking overt verb movement to T. This difference is illustrated in (41) vs. (42):

(41) a. *Það virðist oftast einhver stúdent$_i$ vera t_i í herberginu.* (Ice)
there seems usually some student be in the room
b. *Það virðist oftast vera einhver stúdent í herberginu.*
there seems usually be some student in the room
'There usually seems to be some student in the room.'

(42) a. **Det verkar vanligen någon student$_i$ vara t_i i rummet.* (Swe)
there seems usually some student be in the room
b. *Det verkar vanligen vara någon student i rummet.*
there seems usually be some student in the room
'There usually seems to be some student in the room.'

Again, the claim is that the extra subject position licensed in Icelandic, but not in Swedish, is Spec-TP.[26] But since the indicated movement of the lexical argument in non-finite clauses like (41a) can hardly be for Case checking purposes, under the generally accepted assumption that non-finite verbs cannot check (Nominative) Case on lexical subjects, JONAS argues that the feature checking involved must concern the EPP. Her paper thus contains interesting proposals about the nature of feature checking and its relationship to (overt) verb movement.

One interesting prediction that follows from JONAS' analysis is that if a language has a transitive expletive construction (TEC), it should have overt verb movement (assuming that the associate of the expletive (the logical subject) is in Spec-TP in such constructions). Since Dutch for instance has TEC, it should have overt V-to-T. As is well known, this has been a matter of great debate, with the pendulum swinging towards "no-overt-V-to-T" analyses the past five years or so. Thus Reuland (1990) presents arguments against V-to-T in Dutch, and recent work which assumes that IP and VP are left headed in Dutch (see, for instance, the papers discussed in section 2.6 below and references cited there) also contains arguments to the same effect. JONAS' analysis provides an interesting contribution to this debate.

Finally, it should be mentioned here that at least two of the contributors to this volume suggest that Case checking may not take place exclusively in a Spec-Head relation. One is

[25] For a different account of this extra subject position in languages like Icelandic see Chomsky 1995b:354, where it is argued that the cross-linguistic difference involved is that T in some languages can license two specifiers but only one in others.

[26] THRÁINSSON argues in his contribution to this volume that the crucial difference between languages like Icelandic and Swedish does not involve properties of T as such but rather that Swedish has an unsplit IP (i.e., it does not have AgrS and T as separate functional categories) whereas Icelandic has a split IP, namely both AgrSP and TP. Hence the difference in possible subject positions. He wants to relate this to differences in the morphological inventory of the languages. See the discussion in 2.5 below.

THRÁINSSON, who mentions this possibility in connection with his claim that the Mainland Scandinavian languages (which include Swedish but not Icelandic) do not have an AgrSP as a projection separate from TP, and hence possibly not an AgrOP either if AgrSP and AgrOP are "indivisible" (cf. Chomsky 1993:7). If that is correct, one would seem to be forced to some sort of Case-checking-under-government analysis for Mainland Scandinavian and other languages that do not have an AgrOP. THRÁINSSON does not pursue this point further, however. The other contributor suggesting alternatives to Case checking in a Spec-Head relation is FERGUSON, who suggests that N^{o} object incorporation is in fact a Case feature checking operation on a par with movement of a DP object to a specifier position, accounting for apparent exemptions from the Case Filter in incorporation structures (cf. Baker 1988). But as FERGUSON's paper is mainly concerned with the proper formulation of economy conditions (the concept "Shortest Move"), we leave further discussion of his paper to the next section of this Introduction.

2.3 Economy conditions and restrictions on movement

In section 1.4 we briefly introduced the economy conditions and restrictions on movement of Chomsky (1993). Here, we will discuss these conditions and restrictions a bit more, as they play a major role in most of the articles in this volume.

In Chomsky (1993), economy of derivation is expressed in two requirements, which we will refer to as the shortest steps requirement and the fewest steps (shortest derivation) requirement.

The shortest steps requirement prohibits movement across a potential landing site. It captures the basic phenomena adduced in support of the Relativized Minimality condition on movement (Rizzi 1990).

The fewest steps requirement entails that a derivation should exhibit as few steps as possible. This is expressed in the principle of Greed: superfluous movement is not allowed. Procrastinate, the principle that movement must be postponed until LF, can also be seen as an implementation of the fewest steps requirement.

We will discuss the shortest steps requirement and the fewest steps requirement in 2.3.1 and 2.3.2. Section 2.3.3 addresses the question of whether optional movement should (or can) be allowed in a minimalist approach, and section 2.3.4 discusses another condition on movement, the extension (or strict cycle) condition.

2.3.1 Shortest Steps

The hypothesis that long distance movement proceeds via a succession of local steps is probably one of the more important within generative grammar (Chomsky 1973). Much effort has since been devoted to the exact definition of what constitutes a local domain (see Koster 1978, Chomsky 1981, Kayne 1984, Chomsky 1986a, Rizzi 1990, Cinque 1990, Manzini 1992, and the references cited there).

The concept of minimality, introduced in Chomsky (1986a), relates locality to the domain of a governing head: no head may govern into the domain of another head. Chomsky (1986a) also proposes that head movement enlarges the local domain. This idea re-emerges in the Minimalist Program.

Relativized minimality (Rizzi 1990) introduces another locality condition, one for which the presence of a head is irrelevant. The condition basically says that movement of an element to an α position across another α position is not allowed, where α ranges over {X°, A, A'}. In this approach, nonlocal head movement is not allowed because it skips a head position, nonlocal raising is not allowed because it skips an A-position, and nonlocal Wh-movement is not allowed because it skips an A'-position. This idea also returns in the Minimalist Program.

In Chomsky (1993), the shortest steps requirement forces movement to the closest potential landing site. What constitutes a potential landing site depends on the type of movement, much like in the relativized minimality approach. Thus, for head movement, the closest potential landing site is the first head up, etc.

Assuming now that subjects are generated inside the VP, and that subjects and objects are licensed in the functional domain associated with V (i.e. outside the VP), a problem arises with respect to the movement of the subject and the object out of the VP. The positions in which subject and object are generated are A-positions (see section 2.4). The landing sites for subject movement (Spec-AgrSP or Spec-TP; we will leave this open here) and object movement (Spec-AgrOP) are also A-positions. Movement of the object across Spec-VP therefore violates the shortest steps requirement:

(43) $[_{AgrSP/TP}$ -- AgrS°/T° $[_{AgrOP}$ -- AgrO° $[_{VP}$ SUBJ $[_{V'}$ V° OBJ]]]]

To solve this problem, Chomsky proposes that head movement of V to AgrO enlarges the local domain to which the shortest steps requirement is sensitive (in this case the local domain is VP). Chomsky's formulation of this proposal is that V-to-AgrO movement makes Spec-VP and Spec-AgrOP *equidistant* from the position of the object (OBJ). Since Spec-VP and Spec-AgrOP are equidistant from OBJ, Spec-VP is not a closer landing site than Spec-AgrOP, and movement of the object to Spec-AgrOP therefore satisfies the shortest steps requirement.

Chomsky's solution to the object movement problem now predicts that "overt object raising will be possible only with overt V-raising" (1993:18). That there is a connection between overt verb raising and overt object raising has been suggested by Holmberg (1986). In fact, Chomsky's analysis appears to predict what is known in the literature as *Holmberg's Generalization*. Holmberg's Generalization is based on facts from Swedish and Icelandic, where movement of a pronoun (in Swedish) or noun phrase (in Icelandic) is not possible in past participle constructions. If we assume that the participle does not move, it appears that object movement is dependent on verb movement:

(44) a. *Johan köpte den inte* (Swe)
John bought it not

b. **Johan har den inte köpt*
John has it not bought

c. *Johan har inte köpt den*
John has not bought it

However, it is not clear that Holmberg's Generalization has the scope that Chomsky predicts it to have. For instance, Zwart (1993b, 1994b) argues that in the Continental West Germanic

languages (Dutch, German, Frisian), where overt object raising takes place, the verb does not move to AgrO in embedded clauses (in main clauses, the verb is always fronted):

(45)	a.	*...dat*	*Jan*	*het boek*	*niet*	*koopt*	(Dutch)
		that	John	the book	not	buys	
	b.	**...dat*	*Jan*	*het boek*	*koopt*	*niet*	
		that	John	the book	buys	not	

There seem to be two ways to align the Continental West Germanic facts with the Scandinavian facts. First, we could hypothesize that the object raising in Continental West Germanic is not an object raising of the type under discussion here (i.e., A-movement), but movement to an A'-position. That there are two types of object movement in Dutch and German has been claimed various times in the literature (see Vanden Wyngaerd 1989a, Mahajan 1990). In this volume, DEN DIKKEN presents argumentation supporting the hypothesis that object shift in Dutch may be A'-movement. (FERGUSON addresses the same point in his footnote 11.) But this does not immediately solve the problem, assuming that the object has to undergo A-movement out of the VP before it can undergo the A'-movement. If Chomsky's prediction is correct, the A-movement would have to be accompanied by overt V-raising, contrary to fact (cf. (45b)). Second, one might assume that the Continental West Germanic languages differ from the Scandinavian languages and English in that AgrO is situated to the right in Continental West Germanic, and to the left in Scandinavian and English. This ties in with the traditional analysis of Dutch and German as head-initial languages (Bach 1962, Koster 1975), but is not compatible with Kayne's phrase structure generalizations (see 1.7 and 2.6).

In this volume, Holmberg's Generalization is discussed in the articles by DEN DIKKEN, FERGUSON, GROAT AND O'NEIL, and THRÁINSSON.

DEN DIKKEN stresses the point that in Dutch there are two types of object movement: object raising to Spec-AgrOP, and an A'-movement which he calls ``scrambling''.

FERGUSON proposes a relaxation of the shortest move requirement of Chomsky, essentially refining the notion of "potential landing site". In FERGUSON's article, not every A-position, for example, is a potential landing site for every A-movement. A given position α is a potential landing site for β only if β's features can be checked in α. This solves the object shift problem immediately: Spec-VP is not a position in which the object's features can be checked. FERGUSON points out that his reformulation of the shortest move requirement still makes it possible to derive Holmberg's Generalization. The crucial assumption here (also made in Chomsky 1993:8, and Chomsky 1995b:352) is that the object is not checked by AgrO but by V. An object can be licensed either by incorporation of N° into V (an option not available in Scandinavian languages), or by movement to Spec-AgrOP. But since V has to check the object, the object raising has to be accompanied by verb raising.

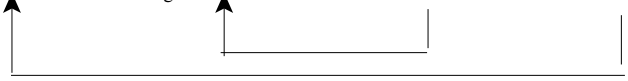
GROAT AND O'NEIL, while accepting Holmberg's Generalization, reject Chomsky's analysis of object shift, in particular the idea that verb movement makes specifier positions equidistant. Instead, GROAT AND O'NEIL link the presence of the object in Spec-AgrOP to the presence in AgrS/T of the verb selecting the object (here, the discussion is restricted to Icelandic, which has fronting of finite verbs in both main and embedded clauses).

Finally, THRÁINSSON discusses Holmberg's Generalization in the light of an analysis that assumes an impoverished functional domain for some languages but not others.

ZWART's article takes an entirely different stand. ZWART takes the incompatibility of the shortest steps requirement and the fewest steps requirement as a starting point, and argues that the shortest steps requirement is superfluous. One of the arguments is that most cases of nonlocal movement that we want to rule out involve movement across a position where certain features must be checked by the moved element.

In defense of the shortest steps requirement, it should be noted here that it is this requirement that derives (part of) the structure of the functional domain. Thus, AgrOP is contained in AgrSP/TP because if AgrSP were contained in AgrOP, the subject would have to move first, and the object would have to cross too many A-positions:

(46) $[_{AgrOP}$ -- AgrO° $[_{AgrSP}$ -- AgrS° $[_{VP}$ SUBJ $[_{V'}$ V° OBJ]]]]



According to the definitions in Chomsky (1993:11f), Spec-AgrOP cannot be made equidistant to both Spec-AgrSP and Spec-VP. This is because verb raising is a stepwise head-adjunction process, involving in this case a V-to-AgrS step and a AgrS-to-AgrO step. Equidistance is calculated after each step, and because of that, the number of equidistant specifier positions is never greater than 2. Thus, if Merge generated the structure in (46), the object could never be checked. The structure in (43) is therefore postulated.

2.3.2 Fewest Steps

In the Principles and Parameters framework (Chomsky 1981), applocation of the transformational rule Move α was unconstrained. Overgeneration was blocked by principles governing the output of movement (such as the Empty Category Principle).[27]

Perhaps the heart of the minimalist movement theory is the hypothesis that movement can apply only if it is triggered by the requirement that some formal feature be checked. Checking results in elimination of the feature. Since the relevant features are there for syntactic purposes only, failure to eliminate them would lead to a violation of Full Interpretation: the output representation of the syntactic component would contain elements that are not interpretable outside the grammatical component.

An even stronger requirement states that α can move only if α carries a feature that is checked (i.e. eliminated) as a result of the movement. This is the principle of Greed (section 1.4).

In Chomsky (1995b), two major changes regarding movement and feature checking are proposed. First, it is proposed that checking does not automatically result in elimination of the feature. Some features, such as categorial features, need to be interpreted at LF, and therefore cannot be eliminated (even if they must be checked). (The wh-features of wh-elements are among these interpretable features, as we have seen above.) Second, the trigger for movement

[27] Nevertheless, it was always clear that some connection between movement and "something else" had to be there, otherwise the movement would be an ad hoc element of the analysis. This "something else" could be formal licensing of noun phrases (Case assignment), overt expression of scope (Wh-movement), linking of a verb to tense (verb raising), etc. In other words, even though movement was unconstrained in principle, there was a strong tendency to avoid arbitrary movement.

has been changed: the trigger used to be an unchecked feature of the element that moves, but now the trigger is an unchecked feature of the head targeted by the element that moves. We could say that a head that has an unchecked feature "looks for" a moveable element that has the corresponding formal feature (e.g., T° looks for a noun phrase with nominative Case features), and *attracts* it. The consequence is that an element that has moved and has its features checked (such as a wh-element that has moved to Spec-CP in the embedded clause) can still move on to check features of some higher head (the matrix C°), provided the relevant features are of the interpretable type and have not been eliminated. In this volume, the articles address the movement theory of Chomsky (1993), which incorporates Greed.

Another principle that could be related to the fewest steps requirement is the Procrastinate principle. Movement is covert, unless the feature triggering movement is "strong". As a result, overt syntax is characterized by a well defined, restricted number of movements, with no arbitrary movements allowed.

As has been noted above, the fewest steps requirement and the shortest steps requirement appear to be incompatible. Chomsky (1993b:15) solves this problem by introducing the operation Form Chain. The articles by ZWART and POOLE are attempts to further define this operation.

ZWART argues that Form Chain should be decomposed into the two familiar structure building operations, Merge and Move. In the particular case of long distance Wh-movement, he assumes that the intermediate trace is in fact an empty wh-element which is merged with the embedded C' (i.e., base generated in the Spec-CP of the embedded clause). The lexical wh-element then moves in one step from its base position to the Spec-CP of the matrix clause. This derivation maximally satisfies the fewest steps requirement, and does not incur the problem that features checked in the intermediate landing site should still be visible to trigger further movement. In this analysis, the lexical wh-element in the matrix Spec-CP, the empty wh-element in the embedded Spec-CP, and the trace of the lexical wh-element are all combined in a chain, which is needed to link the lexical wh-element to its theta-position.

POOLE maintains successive cyclic movement as a succession of movement steps, but makes the Form Chain mechanism more explicit. Form Chain operates on the result of the Move operations, and ``hooks together'' the moved category and its traces. POOLE then proposes that only the Form Chain operation is subject to Greed, while Move is not. This solves the problem of successive cyclic movement containing non-greedy movements (i.e. movement of elements that have their features checked already). Since Move is not subject to Greed, the fewest steps requirement must be thought of as being less crucial to economy of derivation than the shortest steps requirement.

Thus, while ZWART argues against the shortest steps requirement, and wants to give the fewest steps requirement center stage, POOLE argues for the exact opposite.

Greed plays an important role behind the scenes in FERGUSON's article. FERGUSON essentially relates "shortness" to the presence of relevant features (i.e., to Greed). Elements must make the shortest greedy movement, which amounts to saying that the fewest steps requirement is more important than the shortest steps requirement.

The Procrastinate principle is addressed in the article by GROAT AND O'NEIL. Procrastinate expresses an asymmetry between overt and covert syntax, and GROAT AND O'NEIL's article can be seen as an attempt to eliminate such asymmetries. GROAT AND O'NEIL argue that Procrastinate is a "global economy" principle, a principle comparing possible (convergent) derivations. They seek to reformulate Procrastinate as a purely local constraint that operates upon one derivation at a time. The heart of the proposal is that movement

always takes place in overt syntax, but that languages differ as to which of the copies (the head or the tail of the chain) is spelled out. This can be expressed in terms of strong and weak features (as in standard minimalism), but in this approach, we no longer have to say anything about which copy is spelled out in the default case. Procrastinate is precisely such a default rule.

2.3.3 Optionality

It seems that in the strictest implementation of the minimalist framework, optional movement is not allowed. In Chomsky (1991:431), optional movement patterns were considered to be the result of two derivations being equally costly. In the minimalist approach, this would have to be reformulated in terms of features being optionally strong, an unattractive result.

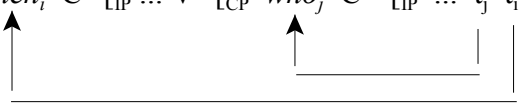
This problem is addressed in POOLE's article. As we have seen in section 2.3.1, POOLE proposes to separate Move and Form Chain, and argues that only Form Chain is subject to economy of derivation. In other words, movement is "cost free". POOLE then proposes that optional movement occurs whenever Move is not accompanied by Form Chain. In those cases, a derivation with movement is not more costly than a derivation without movement. This re-introduces Chomsky's (1991) analysis of optionality as resulting from two derivations that are equally costly.

2.3.4 Strict Cyclicity

The strict cycle condition (Chomsky 1973) was introduced into syntactic theory as a rule ordering constraint needed to account for island effects in successive cyclic movement constructions. Consider the wh-island construction in (47a), schematically represented in (47b):

(47) a. **When$_i$ did you wonder who$_j$ Bill kissed t_j t_i*

b. $[_{CP}$ *when$_i$* C° $[_{IP}$ ... V° $[_{CP}$ *who$_j$* C° $[_{IP}$ ... t_j t_i]]]]



In (47), *who* occupies the intermediate landing site for *when*, blocking successive cyclic movement. The strict cycle condition is needed to ensure that *who* moves into the Spec-CP of the embedded clause before *when* moves. In other words, we must make sure that the island configuration is not created after the movement it is supposed to block has taken place. (See Freidin 1978 for argumentation that the strict cycle condition is superfluous in this domain if we assume that successive cyclic movement leaves traces.)

In the minimalist framework, the strict cycle condition is reintroduced as a condition on the operations Merge and Move, stating that Move can only target the root of a phrase marker. Movement of *who* in (47) can target the root only if it applies before the embedded CP is combined with the matrix V° (hence, also before *when* moves). The strict cycle condition (or "extension condition") ensures that tree structures are built from the bottom up, in an orderly fashion. As illustrated in (47), Chomsky employs the condition in order to

complement his analysis of Relativized Minimality phenomena in terms of the shortest move condition. In addition, the extension condition prohibits movement into complement position (which makes other principles prohibiting this, such as the Theta Criterion and the Projection Principle, superfluous in this domain).

Chomsky notes, however, that the extension condition does not operate without exceptions. Head movement does not seem subject to the extension condition. Chomsky also tentatively restricts the extension condition to substitution (i.e., countercyclic adjunction is allowed) in overt syntax (i.e. covert countercyclic substitution is allowed) (1993:24).

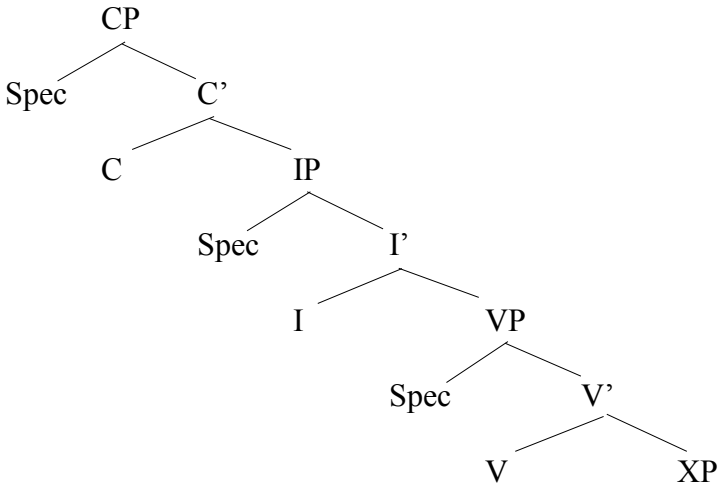
The asymmetry between overt and covert syntax is eliminated in GROAT AND O'NEIL's article. In their analysis, all movement is overt, and language variation is a matter of spelling out the higher or the lower copy (the head or the foot of the chain). Hence, there is no covert substitution (for example, movement of the object to Spec-AgrOP in languages without overt object raising, like English).

The remaining exception to the extension condition, adjunction, does not seem to have any substance outside the domain of head movement. (See section 2.1 for the somewhat tenuous distinction between substitution and adjunction in the minimalist framework.) This suggests that head movement has a special status, and that all XP-movement can be constrained by the extension condition. There have been several recent attempts to derive the special status of head movement (Kitahara 1994, Watanabe 1994, Collins 1994). ZWART's paper also addresses the issue, proposing that the extension condition prohibits projection of an XP inside another XP. This would make countercyclic head movement possible.

2.4 Phrase structure and category types

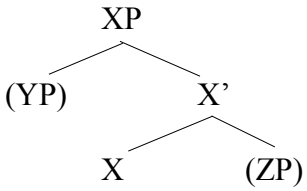
As discussed in sections 1.2 and 1.5 above, the distinction between lexical categories on the one hand and functional categories on the other is a crucial one in most contemporary theories of syntax, although the proposed inventory of functional categories will vary. Consider for instance the structure in (48), which would be typical for, say, Chomsky (1986b) and later work:

(48)



This structure conforms to the X-bar theory outlined in Chomsky (1986b:3), which states that syntactic projections conform to the so-called X-bar-schema illustrated in (49) (leaving aside adjunction):

(49)

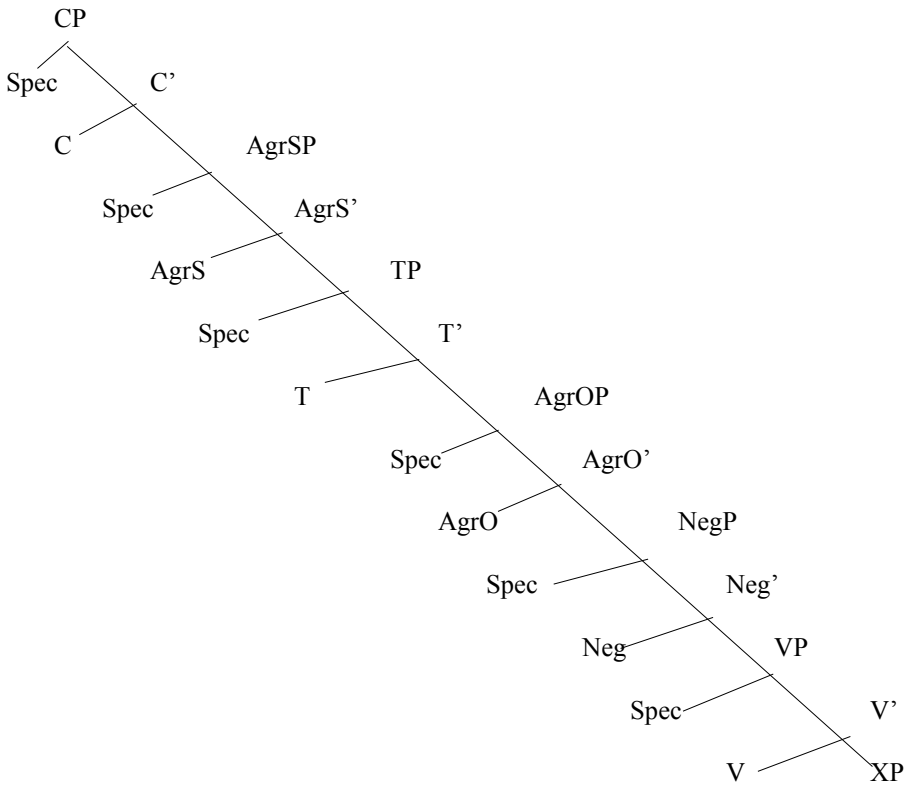


XP in (49) is the maximal projection of the head X, YP is its specifier and ZP its complement. In a structure like (48), the projection VP would be a (maximal) lexical projection of the (lexical) head V, whereas CP and IP are (maximal) functional projections of the (functional) heads C and I. XP in (48) stands for the complement, which could be a maximal projection of various kinds, including CP and NP (or DP).

In the Government-Binding framework, it is assumed that "not all positions are created equal". More specifically, it is assumed that there is an important distinction to be drawn between so-called argument positions (A-positions) on the one hand and non-argument positions (or A'-positions) on the other, where A-positions are those that "may (although they need not) be filled by arguments, the latter being quasireferential elements that require a semantic role" (Chomsky 1986a:80). This means that in (48) the specifier position of IP, the specifier position of VP and the complement position of VP will be the A-positions, if we assume the VP-Internal Subject Hypothesis, according to which Spec-VP is the canonical D-structure (initial) position of the subject, whereas Spec-IP is its characteristic S-structure (Spell-Out) position.

As the distinction between A- and A'-positions plays a crucial role in Government Binding theory, it is important to know how it is translated into the Minimalist Program, which employs different ways of building syntactic structures and also tends to assume a more articulated syntactic structure, sometimes with quite a proliferation of functional categories. Consider in contrast to (48) a clause structure representation like the following, for instance:

(50)



As mentioned in section 1.5 above, Chomsky has suggested that the relevant distinctions are better captured by the notions L-related vs. non-L-related in the Minimalist Program. More specifically, "[a] structural position that is narrowly L-related has the basic properties of A-positions; one that is not L-related has the basic properties of A'-positions" (Chomsky 1993: 28-29), where L-related means 'in a local relation to a morphological feature of a lexical element'. The partition between L-related and non-L-related positions is thus obviously based on a quite different definition than the earlier one between A- and A'-positions. One is based on morphological features, the other mentions elements requiring a semantic (i.e. thematic) role. It is thus an interesting empirical question whether the distinctions made on the basis of these different definitions are empirically equivalent. This is one of the main concerns of HAEGEMAN's paper in this volume.

HAEGEMAN discusses different types of clitics, in both Germanic and Romance, analyzing their similarities and differences. Whereas Romance clitics are argued to be verbal and to interact with verb movement, Germanic clitics are analyzed as non-verbal and move to a functional head that is not L-related in the minimalist sense. HAEGEMAN provides an overview of some of the discussion of the L-related/non-L-related typology of syntactic positions within the Minimalist Program and relates this to her analysis, with the main emphasis on the behavior of West Flemish clitics. She claims that although the distinction between L-related and non-L-related positions is a linguistically significant one, it does not in fact replace the A- vs. A'-distinction. In particular, she maintains that there appear to exist A-positions that are not L-related. This is clearly a central and hence important claim regarding the typology of syntactic positions specified in UG. It is well known that a simple A/A' distinction seems insufficient in many cases, e.g. in connection with the analysis of the properties of Scrambling (see e.g. the discussions in Webelhuth 1992 and Mahajan 1990) and

it would be interesting to see whether HAEGEMAN's proposal can help shed new light on some old A/A' problems.

A related issue concerns similarities and differences between specifiers and adjoined phrases. Chomsky (1993:28) states that the specifier of a head with an L-feature (a morphological feature) is narrowly L-related whereas a category adjoined to the maximal projection of such a head would be broadly L-related. As we saw above, Chomsky suggests that only the first type (i.e. the narrowly L-related position) has the properties of an A-position. This is quite standardly assumed in much recent syntactic work. The nature of the whole issue becomes somewhat different in a framework where multiple specifiers of a single head are allowed, as assumed, for instance, in some of Chomsky's most recent work (see Chomsky 1995b and references cited there). The multiple specifier analysis provides an alternative to the AgrSP+TP analysis advocated in the papers by JONAS and THRÁINSSON in the present volume in connection with transitive expletive constructions and other "multiple subject" constructions (see the discussions in section 2.2 above and 2.5 below). But, as pointed out by these authors, it is not immediately clear what the multiple specifier analysis entails regarding the syntax of the V2 phenomenon as it appears in transitive expletive constructions, for instance. The reason is that if we assume that the expletive occurs in the higher specifier and its associate in the lower specifier of the same projection (say TP) in such constructions, the appearance of the verb between the two rather than in the head position "below" both (i.e. in T in this case) has no obvious syntactic analysis. Hence the multiple specifier analysis is forced to postulate that the "second position" of the verb in such instances is a ``phonological" phenomenon.

Finally, as a comparison of the structures in (48) and (50) above suggests, work in the Minimalist Program tends to assume a large number of functional categories. The suggestion that the structure of the Inflection Phrase (IP) is more complex than illustrated in (48) and that it should be decomposed into an Agreement Phrase (AgrP) and a Tense Phrase (TP) goes back to Pollock's influential (1989) paper, thus predating minimalism. But functional categories play an even more important role in the Minimalist Program than before. Consequently it is important to ask about their precise role(s) and universality. Do different clause types, for instance, vary with respect to which functional projections they instantiate? Although it sometimes seems to be assumed that the functional categories present in all types of clauses are invariant, it should be pointed out that the common hypothesis that Small Clauses exist (i.e. clauses without an I-projection, see e.g. Stowell 1983), presupposes that clause types may indeed vary in terms of the functional categories they contain. This question is discussed in some detail in THRÁINSSON's paper in this volume, as will become clear in the following section.

2.5 Universality, parametric variation and morphological strength

2.5.1 Some basic ideas

As discussed in sections 1.1 and 1.6 above, two of the major concerns of modern linguistic theory are arguably the following:

(51) a. What is universally true of all grammars of natural languages - and by hypothesis, known by humans in advance of any linguistic experience?

b. To what extent can human grammars vary and how does the "learner" come to know the particular variant commonly referred to as "the language the child has learned"?

This emphasis on discovering what is universal and what is not is particularly apparent in the so-called Principles-and-Parameters approach to syntax. The origin and nature of the basic ideas underlying this approach is described in a very accessible fashion in Williams (1987) and Freidin (1991), for instance, and the approach is further explicated by Chomsky and Lasnik (1993).

An important contribution to this discussion of universality and linguistic variation is Pollock's seminal (1989) paper on verb movement and clause structure. As is well known, Pollock's main innovation was to suggest that it is possible to account for certain cross-linguistic differences if one assumes that the categorial inventory includes an Agreement (Agr) projection and a Tense (T) projection rather than just a single Inflection (I) projection as had been previously assumed. Pollock further suggested that the Agr-projection could have different properties cross-linguistically, depending on whether the language in question had a "morphologically "rich" system of agreement inflections" or not (see, e.g., Pollock 1989: 418-419). As Pollock points out, Roberts (1985) had also attempted to explain cross-linguistic syntactic differences in terms of varying degrees of richness of inflection but without assuming a separate Agr-projection. In further development of Pollock's approach to clause structure, Chomsky (1991) and Belletti (1990) also discuss the notion of "morphological strength/ richness" employed by Pollock. Inspired by research of this kind, many syntacticians investigating comparative Germanic and Romance syntax have tried to determine how "rich" verbal morphology must be in order to make the relevant inflectional projections "strong". More specifically, what these authors are mainly concerned with is the fact that there seems to be evidence for overt verb movement in some languages but not others and the presence of overt verb movement seems to have some correlation with richness of inflection. More recent research on comparative Germanic and Romance syntax has sought to define the relevant notion of "rich inflection", including work by Roberts (1993), Rohrbacher (1994), Vikner (1995a, 1995b), and others. We shall return to this question below.

Another attempt to account for the relationship between morphology and syntax is embodied in Baker's Mirror Principle (see e.g. Baker 1985, 1988), which maintains that there is a direct relationship between syntactic derivations (and thus syntactic structure) and (overt) morphology. In a framework where it is assumed that inflectional morphemes, such as agreement and tense morphemes, are functional heads and these inflectional elements are then "picked up" by lexical heads that adjoin to (unite with) the relevant functional head positions, the implementation of the basic idea behind the Mirror Principle is relatively straightforward. It is not as clear, however, what role the Mirror Principle can play in the strictly lexicalist approach embodied in the Minimalist Program.[28]

[28] This simplified exposition ignores the fact that languages vary with respect to "overt verb movement". As discussed by Pollock (1989) and Chomsky (1991), for instance, French differs from English in this respect. Hence the idea (which goes back to Chomsky 1957) that a rule of Affix Hopping (or Affix Lowering) might be an appropriate way of accounting for aspects of the morphosyntax of finite verbs in English while a rule of Verb Raising (overt verb movement that is) would be appropriate for French. The strong lexicalist approach to morphosyntax adopted in the Minimalist Program (Chomsky 1993) makes it possible to minimize this difference. Under such an

These are some of the basic issues having to do with the universality/diversity of functional categories and their relation to morphology. Some of the papers in this volume discuss these issues in considerable detail and the following two subsections contain a brief introduction to these discussions.

2.5.2 Functional categories, parametric variation, and overt morphology

Chomsky's original work on the Minimalist Program (1993) assumes the so-called "Split Infl Hypothesis" of Pollock (1989) which had been further developed in Chomsky's own work (1991) and by Belletti (1990) and others. However, instead of just one Agr-projection, as Pollock had originally suggested, Chomsky (1993), following Chomsky (1991:434), postulates an AgrS-projection that enters into a checking relationship (Spec-Head relationship) with the subject, and an AgrO projection that plays a corresponding role in object checking. Both Chomsky (1991) and Belletti (1990) argue that AgrSP dominates TP and this seems to be quite generally accepted in the pre-minimalist literature that adopts the Split Infl Hypothesis. Work by Ouhalla (1988, 1991, 1994) and Campbell (1991) is somewhat of an exception in this respect, as they have argued that the hierarchical order of AgrSP and TP may vary and that this variation reflects differences in morphological structure according to Baker's Mirror Principle. Thus if the subject agreement morpheme is "further away from" the stem of the verb than the tense morpheme is, then the syntactic AgrS projection (AgrSP) will dominate the T projection (TP).[29]

In the Minimalist Program, a strong lexicalist approach to inflectional morphology is postulated in the sense that words are supposed to emerge from the Lexicon fully inflected. Thus there is no affixation of inflectional morphemes via the application of syntactic head-adjunction. Instead, it is assumed that morphological features of lexical elements are checked against matching features of functional heads in the syntactic structure. As pointed out above, it is not immediately obvious how the Mirror Principle could be cast within such a model. In his Minimalist Program paper, Chomsky (1993: 28) makes a brief suggestion about how the spirit of the Mirror Principle could be maintained under the strong lexicalist approach he assumes. This idea is taken up in THRÁINSSON's paper in the present volume and developed further. He suggests that if we assume that morphological features are associated with particular inflectional morphemes (to the extent that such morphemes can be found in the language in question) and if the features of morphemes closer to the stem must be checked first, then we can capture the essence of the Mirror Principle in a lexicalist framework (see Halle and Marantz 1993:166f for discussion)..

Now it should be noted that the relationship between morphological feature checking and overt morphology assumed by THRÁINSSON is much closer than assumed in Chomsky's Minimalist Program paper (1993). Thus while Chomsky keeps the basic idea advocated by

approach the difference between French and English is no longer claimed to be that the former has Verb Raising and the latter Affix Lowering but rather that the former has overt Verb Raising and the latter covert Verb Raising.

[29] Ouhalla argues that while the agreement marker is further from the stem of the verb than is the tense marker in French and thus AgrSP dominates TP, the agreement marker is closer to the stem of the verb than is the tense marker in Arabic and hence the T projection dominates the AgrS projection (see e.g. Ouhalla 1994: 45. - For a different account see Shlonsky, in press.). Campbell (1991) argues that within a single language, German, TP dominates AgrSP in the present tense, whereas in the past tense, AgrSP dominates TP.

Pollock and many others that overt verb movement is triggered by "strong morphology" (see the references in 2.5.1 above), he argues that the parametric variation should be expressed by assuming strong vs. weak features on functional heads and that these features are abstract in the sense that their value has no direct relationship to overt morphological distinctions. SOLÀ's paper in the present volume is very similar in spirit to THRÁINSSON's. Thus SOLÀ argues that the crucial parameter which determines the occurrence of overt head movement in general (and overt verb movement in particular) is "the presence/absence of inflectional morphology for a given functional category on a given word". We will return to this aspect of SOLÀ's paper in section 2.5.3 below.

Attempts to relate functional structure in the syntax to overt morphological structure in such a direct fashion raise the question of the universality of functional categories. As it is well known that languages exhibit considerable variation in overt inflectional morphology, directly relating functional structure to inflectional morphology would seem to predict a great cross-linguistic variation in functional structure. Such a concept of functional structure seems clearly at odds with much recent work in theoretical syntax, which tends to assume (near) universality of functional structures, as THRÁINSSON notes in his present paper. But Pollock's and Chomsky's original ideas about the decomposition of the Infl-projection were actually criticized early on by Iatridou (1990). Iatridou warned against the idea that all languages should be analyzed as having Agr- and T-projections, for instance, although there might be (morphological and syntactic) evidence for them in some languages. THRÁINSSON's paper is close in spirit to Iatridou's, as one of his main claims is that only some languages have a split Infl in the Pollockian sense whereas others simply have a single unsplit IP-projection. THRÁINSSON goes on to discuss how morphological triggers may facilitate acquisition of the value of what he refers to as "The Split IP Parameter". He claims, for instance, that English and Mainland Scandinavian (MSc) have a "fused" IP whereas French and Icelandic have a split IP. This is supposed to explain some syntactic differences between these two types of languages, including the availability of the "extra subject position" in Icelandic as opposed to MSc. THRÁINSSON agrees with JONAS that this extra subject position is Spec-TP and argues that the reason it is not available in MSc is that MSc simply does not have a T-projection separate from the AgrS-projection and hence there is only one such specifier (subject) position present, namely Spec-IP. This is then related to the fact that MSc does not have separate tense and agreement morphemes whereas Icelandic does.[30]

Now, if THRÁINSSON and SOLÀ are right in that there is a rather direct relationship between functional structure and syntactic movement on the one hand and overt morphology on the other, then the question arises whether different clause types within a given language vary with respect to the functional categories they instantiate. If the Agr-projection, for instance, is really an agreement projection, then we might expect it to be absent in non-finite clauses since these typically do not exhibit (overt) agreement. This issue is briefly addressed in THRÁINSSON's paper and also in the paper by JONAS, but it is also related to the question of

[30] As discussed in THRÁINSSON's paper, a somewhat similar approach is taken by Bobaljik 1995, although Bobaljik does not want to tie the presence of a separate Agr-projection as directly to overt morphology as THRÁINSSON attempts to do. A different position is taken by Chomsky (1995b:354), who suggests that there may be no need for a separate Agr-projection at all in any language and the "extra subject position phenomenon" may instead result from the ability of T to project more than one specifier in some languages but not in others. The basic idea of the relevance of "strong inflection" could be implemented in such an approach by saying that only "strong" T can have this "multiple specifier" property.

what triggers overt movement, namely whether there are abstract strong features or whether overt inflectional morphology is solely responsible. This issue will be examined further in the next subsection.

2.5.3 Morphological strength and overt order

As mentioned above, Chomsky assumes in his Minimalist Program paper (1993) that (abstract) strong features on functional heads trigger overt syntactic movement. The original idea was that strong features had to be checked before Spell-Out since unchecked strong features would be visible but uninterpretable at PF and thus cause the derivation to crash.[31] Weak features, on the other hand, need not be checked in overt syntax, and hence they will not be, since covert movement is more economical than overt movement (cf. the discussion of economy principles in sections 1.4 and 2.3 above).

As already mentioned, SOLÀ rejects the notion of (abstract) strong features and suggests instead that the key to the relevant syntactic variation can be found in overt morphological distinctions. He argues that given the proper morphological analysis, we can account for cross-linguistic differences in functional structure and overt syntactic movement. Most of his arguments are based on comparisons of English and (other) West Germanic languages. His analysis adopts Kayne's (1994) proposals, in particular their adaptation to West Germanic languages along the lines suggested by Zwart (1993b), for instance. This approach delimits cross-linguistic phrase structure variation (see the discussion of directionality in 1.7 above and 2.6 below). The basic idea is that an inflected lexical element moves only if it "contains overt morphology of the target of [the] movement;[32] and there is no covert movement". This is obviously a quite radical departure from the basic idea of Chomsky's Minimalist Program, which assumes both overt and covert movement.

The paper by Groat and O'Neil proposes ideas somewhat similar to SOLÀ's. More specifically, Groat and O'Neil suggest that there are "no post-Spell-Out syntactic operations". In this respect they agree with SOLÀ. Their particular implementation of the idea is different, though. In particular, they actually do assume both overt and covert movement. The difference is not, however, that covert movement applies "after Spell-Out" or "in LF", as standardly assumed in the Minimalist Program, but rather that all movement precedes Spell-

[31] The account of the relationship between strong features and overt movement is somewhat different in Chomsky's most recent work (1995b), but a detailed description of that alternative is not directly relevant here (see also section 2.3.2).

[32] An interesting question here is whether this implies, say, that verb forms not overtly inflected for tense, like infinitives in Germanic languages, would not be expected to move to T. As has been argued, there appears to be verb movement in control infinitives in Icelandic, for instance (see e.g. the papers by JONAS and THRÁINSSON in this volume and references cited there). This could be taken to suggest that although overt morphology may provide clues to the language acquirer about the nature of functional projections in the target language, their "strength" is determined once and for all in each language and does not vary depending on the inflectional properties of the lexical heads in different clause types (finite and non-finite). This is obviously an interesting and broad topic requiring further research.

Out and "overt movement is movement which carries phonological features to the head of the chain being created, while covert movement leaves phonological features behind".[33]

This idea of a single-interface model proposed by Groat and O'Neil will be discussed further in section 2.7 below. It should be noted here, however, that Groat and O'Neil's analysis differs from SOLÀ's in that the former assume that it is feature strength of functional heads that triggers overt movement, rather than overt inflectional morphology of lexical items, as SOLÀ argues.

Finally, an interesting claim made by Groat and O'Neil is that the chains created by overt vs. covert movement have different licensing properties. This is an issue also discussed in the paper by JONAS in the present volume. Recall that JONAS argues that overt verb movement to T licenses Spec-TP as a subject position. Jonas and Bobaljik (1993) had previously argued that having Spec-TP as an available subject position was a prerequisite for object shift in languages like Icelandic, for instance. Groat and O'Neil argue, on the other hand, that overt verb raising to T (or T/AgrS in their terminology) makes Spec-AgrOP an available theta position and thus makes it possible to base-generate the object in Spec-AgrOP, creating the illusion of object shift.

2.6 Directionality and word order

In traditional grammar, in typological research, and in the pre-minimalist stages of generative grammar, a distinction is made between head-final and head-initial languages. It is not entirely correct to equate head final with OV and head initial with VO. A language can be characterized by OV word order without necessarily having head-final structure, if the OV order is the result of object raising. Thus, the VO/OV distinction refers to surface structure, not necessarily to the order of elements in the initial representation (the product of Merge). We can therefore continue to speak of VO, OV, SVO, VSO, etc. languages, without committing ourselves to claims about the basic structure of these languages.

In the minimalist framework, the VO, OV, etc. patterns can be described in terms of the interaction of overt and covert movement. Thus, a language can be characterized as VO if its basic structure is head initial and no overt movements take place, or if the object and the verb move overtly, with the verb ending up to the left of the object, etc. Notice that if we assume that all languages are head initial (or head final, for that matter), the interaction of covert and overt movement is the only factor bringing about the surface OV, VO, etc. distinctions. This is one of the conceptual arguments in Kayne (1994) for assuming that all languages are head initial (see section 1.7).

In the Principles and Parameters framework, it was assumed that the order of head and complement at D-structure was an instance of parametric variation. A language could have one or the other setting of the *headedness parameter*. Kayne (1984) and Koster (1987), among others, derived the headedness parameter from directionality of government: in some

[33] Interestingly, the idea of moving the relevant features and "leaving phonological material behind" is also proposed in Kitahara (1994) and further developed in Chomsky's most recent work (Chomsky 1995b:261ff). The main difference is that Chomsky assumes that this feature movement takes place "after Spell-Out", i.e. at LF, and movement which has overt consequences (pied-pipes phonological material) takes place "before Spell-Out". Groat and O'Neil, on the other hand, assume that Spell-Out is always "applied to the same phrase-marker that receives LF-interpretation".

languages, the verb governs to the right, in others, to the left. If theta-roles are assigned under government, rightward governing verbs take their complement to the right (yielding head initial structures), and leftward governing verbs take their complement to the left (yielding head final structures).

This derivation of the headedness parameter is no longer possible in the minimalist framework, in which the government relation is no longer available. To review briefly, government is not needed for theta-role assignment which is now a function of the operation Merge (i.e. of the sisterhood relation), nor is it needed for Case assignment, which is now redefined as feature checking in a Spec-Head configuration, nor for the formulation of the PRO-theorem, since PRO is now defined as an element with Null Case, not as an ungoverned element (Chomsky and Lasnik 1993), etc.

Viewed from this perspective, Kayne's conjecture that all languages are head-initial is highly consistent with the minimalist framework of Chomsky (1993).[34]

In this volume, Kayne's generalizations regarding phrase structure are adopted and employed in an analysis of Dutch past participle constructions by VANDEN WYNGAERD (see the discussion in 2.1). VANDEN WYNGAERD adopts the analysis of Dutch, an OV language traditionally regarded as head-final (cf. Koster 1975), as involving head-initial structures only (cf. Zwart 1994a; see Zwart, to appear, for an introductory text on the analysis of Dutch as a head-initial OV language).

The other article concentrating on Dutch, by DEN DIKKEN, also adopts Kayne's assumptions, taking Dutch to be a head-initial language. DEN DIKKEN pursues earlier suggestions by Kaan (1992) and Zwart (1993b) and analyzes West Flemish Verb Projection Raising constructions as involving head-initial structures with multiple VPs, and AgrOPs either dominating the VPs or inserted between two VPs. In Verb Projection Raising constructions, like (52), a sentence final cluster of verbs is broken up by material appearing in between the verbs:

(52) *..da Jan wilt geen vlees eten* (West Flemish)
that John wants no meat eat-INF
`...that John does not want to eat meat.'

In the traditional analysis of Dutch as a head-final language, (52) is derived by moving the verb projection *geen vlees eten* 'eat no meat' to the right, across the higher verb *wilt* 'wants'. DEN DIKKEN points out various inadequacies of this traditional approach, and strengthens the Kaynean analysis by deriving certain scope ambiguities (or the absence thereof) in these Verb Projection Raising constructions. In doing so, DEN DIKKEN employs KITAHARA's analysis of scope ambiguities, also presented in this volume.

The analysis of Dutch and German as head-initial languages is not uncontroversial (see e.g. Abraham 1994 and Sternefeld 1994). In this volume, the issue is relevant to the status of Holmberg's Generalization and to the hypothesis advanced in the article by JONAS regarding the distribution of transitive expletive constructions. Holmberg's Generalization correlates overt object raising and overt verb raising. Assuming that Dutch and German have object raising of the relevant type (i.e. movement to Spec-AgrOP), the hypothesis that Dutch and German are head-initial is not consistent with Holmberg's Generalization: the nonadjacency

[34] In Chomsky (1995b:335f), the LCA is adopted as a linearization device operating between Spell-Out and PF. We will not discuss this development here.

of the object and the verb indicates that the verb has not raised to AgrO (see section 2.3.1 above). JONAS presents the hypothesis that languages that have transitive expletive constructions also have overt verb raising to T. But if Dutch, a language with transitive expletive constructions (Bennis 1986, Zwart 1992a), has a head-initial TP, the verb should not be allowed to stay to the far right in embedded transitive expletive constructions, contrary to fact:

(53) *..dat* *er* *veel mensen* *een praatje* *hielden* (Dutch)
that there many people a talk held
'...that many people gave a talk.'

Thus it seems that something has to give here.

2.7 The organization of the grammar

As discussed above, the Minimalist model of grammar eliminates both D-structure (a level which is presumed to interact with no non-linguistic system) and S-structure, a level which similarly interacts with no non-linguistic system while concomitantly interacting with D-structure, PF and LF (in the standard Government and Binding model of Chomsky 1981). As Chomsky and Lasnik (1993:511) note, we might think of S-structure, informally, as a "... (presumably unique) "solution" ...", i.e., the only possible representation "successfully" mediating between D-structure, PF and LF via transformational application. Thus, in Chomsky (1993) S-structure as a level of representation is eliminated—in its place, there is an optional rule of Spell-Out: at any point in a derivation (i.e., after any given transformational rule application) the representation derived may be spelled-out. Doing so sends that representation into two separate, non-interacting components of the grammar: the PF component and the LF (or covert) component. The leading idea is that properties of S-structure representation may well be deducible. For example, if, in an English derivation, the following representation were spelled-out

(54) [*I wonder* [$_{C^0}$ [*John likes what*]]]
+wh
+strong

the strong +wh feature of C° would, as a result, appear in PF. However, strong features are illegitimate PF objects, and consequently the derivation crashes at this level. Thus, a PF (not an S-structure) condition is violated—an empirical hypothesis—carrying with it the intuitively appealing and seemingly correct prediction that (54) is not a correct representation of sound/pronunciation.

While S-structure as a level of representation is targeted for deduction (an appealing goal), notice that in certain central respects the Government and Binding model nonetheless persists—even in the wake of D-structure and S-structure elimination. That is, Spell-Out splits the derivation into two tracks, PF and LF.

Just as van Riemsdijk and Williams (1981) investigated alternatives to the Government and Binding model of Chomsky (1981) (going back to Chomsky and Lasnik 1977), GROAT AND O'NEIL (this volume) explore an alternative to the Minimalist model of grammar. As

GROAT AND O'NEIL note, the Minimalist model incorporates three quite fundamental, but stipulated, asymmetries concerning the application of pre- vs. post-Spell-Out operations:

(55) a. Pre-Spell-Out operations are more costly than are post-Spell-Out operations (Procrastinate) (Chomsky 1993:30).
b. The strict cycle condition (the extension condition) applies only to (certain) pre-Spell-Out operations, but does not constrain post-Spell-Out operations. In particular, LF object-shift in languages like English is apparently countercyclic (non-extending) and hence the extension condition is presumed inapplicable in the covert component (Chomsky 1993:24).
c. There is lexical access before Spell-Out, but not after Spell-Out (Chomsky 1993:22).

As GROAT AND O'NEIL note, the formal operations (binary and singulary Transformations, Merge and Move) which apply before Spell-Out are, by hypothesis, formally identical to those applying in the covert component. Thus, given the three asymmetries just noted, the model of the Minimalist Program in fact postulates two distinct computational systems that are, except for these asymmetries, identical. A worthy goal and the one sought by GROAT AND O'NEIL is the elimination of these asymmetries, yielding a single computational system.

In order to eliminate pre- vs. post-Spell-Out asymmetries, GROAT AND O'NEIL hypothesize a model in which it is logically impossible to even state such asymmetries. In their model, a derivation (iterative transformational-rule application) yields a single phrase-marker which feeds both the phonological and the interpretive components—thus their title "Spell-Out at the LF interface". Given such a model of grammar, there are simply no post-Spell-Out operations and therefore it is impossible to stipulate pre- vs. post-Spell-Out asymmetries. But if one and the same phrase-marker is both spelled-out and undergoes interpretation, how can overt operations (triggered by strong features) vs. covert operations (triggered by weak features) be distinguished?

For GROAT AND O'NEIL, overt vs. covert movement does not concern the derivational timing of movement (pre- vs. post-Spell-Out). Rather, strong feature checking is phonological feature checking, and it therefore requires that movement carry-along the phonological features of the moved category. By contrast, weak feature checking is not phonological feature checking, and it therefore allows phonological features to be left on the trace, i.e., not carried along by movement. Procrastinate (in Chomsky 1993, a global transderivational economy condition) is recast as a local, intraderivational economy condition: "It is more work to carry along phonological features then it is not to: Weak feature checking therefore precludes moving phonological features." (See also Kitahara 1994, 1995, 1996, and Chomsky 1995b:262.) Notice now that English object shift need not be formulated as a countercyclic LF operation. Instead English object shift is cyclic but since it is a weak (N-)feature checking operation, the now local intraderivational economy condition Procrastinate prohibits moving the phonological features of the shifted English object. Given this analysis, the extension condition (strict cycle condition) need not be stipulated to be inapplicable in LF—i.e., this unexplained pre- vs. post-Spell-Out asymmetry is eliminated.

Further support for their model is provided by GROAT AND O'NEIL's re-analysis of subjacency effects in Japanese Wh-movement (Watanabe 1992) and Icelandic object shift.