Principles & Parameters Theory and Minimalism

Željko Bošković
University of Connecticut

The central problem for the linguistic theory, often referred to as Plato's problem, is how children are able to acquire language despite the impoverished nature of the linguistic data they are exposed to. The Principles and Parameters approach provides a viable solution to Plato's problem. The basic idea is that children are biologically endowed with a set of principles, which are invariant across languages, and a set of parameters, which provide limited options for language variation and which can be easily set on the basis of the linguistic data children are exposed to. These principles and parameters comprise what is referred to as Universal Grammar, a genetic endowment which helps children acquire language and as such represents an answer to Plato's problem in the domain of language.

Principles are not subject to language variation, they hold across all languages with equal force. As an illustration, the Theta Criterion (which may be at least partially subsumable under a broader Principle of Full Interpretation, see below) is responsible for the ungrammaticality of *John is likely that Mary is sleeping*. The example contains an argument, namely *John*, which does not bear a theta role, in violation of the Theta Criterion, which requires that all arguments bear theta roles (as well as that all theta roles be assigned.) By providing options parameters account for a good amount of language variation. The head parameter is one such parameter, posited early on in the Government and Binding Theory. According to this parameter, languages differ regarding whether their heads precede or follow complements. Thus, in English both verbs and prepositions precede their complements, while in Japanese they follow them (but see the discussion of Kayne 1994 below). Borer (1984), however, restricts the notion of parameters by assuming that all parameters are lexical, which reduces parametric variation to differences among lexical items, a further restriction being that the variation is tied only to differences among functional elements (i.e. grammatical elements like inflection).

1.Government and Binding Theory

Government and Binding theory (GB) is the version of the Principles and Parameters approach that was dominant in the 80s. GB is a modular theory which divides grammar into a number of distinct subcomponents with a powerful transformational component based on the rule Move Alpha, which in principle allows any element to move anywhere at any point. Overgeneration is handled by having various modules filter out the undesired structures. Four levels of representation are posited where conditions of UG apply, filtering out the illicit structures: D-structure (DS), S-structure (SS), Logical Form (LF), and Phonological Form (PF). DS, SS, and LF together constitute the syntactic levels of representation. The central grammatical relation is government, a powerful grammar-internal (in a sense that it is not determined by anything outside of the grammar) relation that crucially holds in a number of otherwise distinct modules.

(1)
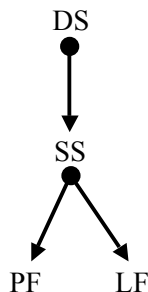
As an illustration, consider the working of the Case module, which determines the distribution of NPs. There are characteristic structural positions that are tied to particular cases; thus, the subject position of a finite clause licenses nominative and the object position accusative case, as in *He likes her*. (The relevant distinctions are overtly manifested in English only with pronouns, but in other languages, e.g. Russian, all

nominals show them.) A central tenet of GB is that NPs (even those that do not overtly show case distinctions) can occur only in positions where case is assigned due to the Case Filter, which filters out all constructions that contain caseless NPs at SS. In a passive like *was arrested Mary, Mary* is base-generated in the object position of the verb to satisfy its thematic properties. However, since, in contrast to active verbs, passive verbs do not assign case, if *Mary* remains in this position the structure will violate the Case Filter. Another relevant condition here is the Extended Projection Principle (EPP), which requires all clauses to have a subject (see chapter 11). One way to satisfy both the EPP and the Case Filter in the construction under consideration is to move *Mary* to the subject position, as in *Mary was arrested.* Alternatively, the EPP can be satisfied by inserting an expletive, as in **it was arrested Mary*. But then *Mary* needs to move to another position to avoid violating the Case Filter. In **Mary seems that it was arrested*, *Mary* is placed in the right position as far as the Case Filter is concerned. However, the movement violates other modules, in particular the locality module. One relevant constraint is Rizzi's (1990) relativized minimality, which prohibits movement to SpecIP (i.e. subject position) across another SpecIP. (More generally, relativized minimality prohibts movement to a position of type X across an element in X, see chapter 17.) The above discussion illustrates how Move Alpha and various conditions and modules that constrain its application work.

GB posits a distinction between overt movement, which takes place before SS (the point when the structure is sent to PF, which determines pronunciation) hence has an effect on pronunciation, and covert movement, which takes place during the mapping from SS to LF, hence does not affect pronunciation. One argument for covert movement involves locality constraints on movement. Thus, (2) is ruled out on the reading on which *how* modifies the embedded clause (which means *how* should be generated within the embedded clause on the reading in question, in the position indicated by t(race)) because movement of *how* from the embedded clause to the matrix SpecCP (the position for fronted wh-phrases in questions) yields a locality violation (Rizzi's relativized minimality is also relevant here, since the example involves movement to SpecCP across a filled SpecCP). Significantly, Chinese (3), where all wh-phrases stay in situ (Chinese being a wh-in-situ language), also lacks the reading on which *how* is interpreted in the matrix clause and *why* in the embedded clause.

(2)   *How do you wonder [why Mary left t]
(3)   *Ni xiangzhidao [Mali weisheme zenmeyang likai]
          you wonder   Mary why   how   left

The unacceptability of the Chinese example on this reading can be accounted for in the same way as the unacceptability of its English counterpart if the Chinese example involves the same movements as the English example, the only difference being that in the Chinese example the movements take place in LF, hence are not reflected in the pronunciation. Under this view English and Chinese differ only in the timing of wh-movement.

2. Minimalist Program

The most recent outgrowth of the Principles and Parameters approach is the Minimalist Program (MP). MP explores the possibility that the content attributed to UG is an optimal way of satisfying requirements imposed on the language faculty by the external systems that the language faculty interfaces with and is characterized by optimal, computationally efficient design. Language is assumed to consist of a lexicon and a computational system. The computational system is embedded in two performance systems: articulatory-perceptual and conceptual-intentional. Two linguistic levels, namely, Phonological Form (PF) and Logical Form (LF), are postulated as interfaces with the performance systems. Each linguistic expression, generated by the computational system, is a pair of representations at PF and LF. A computation converges at the interface levels if its structural description contains only legitimate PF and LF objects, with all of their morphological features satisfied, otherwise, it crashes. However, defining linguistic expressions simply as pairs (P, L) formed by a convergent derivation and satisfying interface conditions does not suffice; the operations of the computational system that produce linguistic expressions must be optimal, in the sense that they must satisfy some general considerations of simplicity

and efficient design. These considerations ban superfluous steps in derivations and superfluous symbols in representations, prefer short movements over long movements, etc. It is then no surprise that the drive for simple and nonredundant explanations (Occam's razor), a natural aspect of any scientific inquiry which has characterized the development of linguistic theory even before minimalism (though probably not as much as the core sciences), has come to the forefront of linguistic theorizing with MP, prompting re-evaluations of the standard tenets of GB from this point of view.

MP can in fact be best understood in relation to its immediate predecessor, GB, a task to which I turn now with respect to some of the most important concepts of MP.

2.1. Syntactic interfaces and levels of representations

Under the traditional view that human language is a way of relating sound and meaning, the interface levels of PF and LF seem to be uneliminable. Minimalism begins with the position that these levels of representation, which are the only conceptually necessary linguistic levels, are in fact the only levels of representation. Elimination of DS and SS has led to a number of fruitful lines of research, including the exploration of the possibility of movement into theta-positions and reanalyses of DS/SS conditions as either PF/LF or derivational conditions. Consider the former (see section 2.4.1. for the latter). DS is an internal interface level with the lexicon which comes with several assumptions regarding lexical insertion that are not conceptually necessary. DS is built by an all-at-once application of an operation that selects an array of lexical items and presents them in a way that satisfies X-bar theory. Since DS is the starting point of the derivation, all lexical insertion then takes place before movement applies. But why would all lexical items have to be introduced all at once, before any movement takes place? MP abandons the 'all at once' and 'before any movement' assumptions. Instead, elements taken from the lexicon project separate subtrees which are combined to form a phrase marker by generalized transformations. Thus, in *these smart women like Mary*, *these smart women* is assembled separately and then combined with *like Mary*, which is also assembled separately. Under this approach lexical insertion can take place after movement. In fact, given the Cycle, which requires syntactic operations to apply at the top of the tree, in *They think that Mary was arrested, Mary* is generated in the object position of *arrest* and then moves to SpecIP of its clause before the lexical items *they* and *think* are introduced into the structure. But if Move can take place before lexical insertion, why couldn't movement then take place into a theta position? Such movement is impossible in a system assuming DS. DS represents the thematic structure of the sentence as a consequence of the Projection Principle, which requires that lexical properties be satisfied, and the Theta Criterion, which requires that every theta role be assigned to an argument and that every argument bear a theta role, applying at DS. Since no theta positions can be left empty at DS, movement into a theta position is not possible. In a system where DS and SS are eliminated, with the Projection Principle and the Theta Criterion holding trivially at LF (if they are not satisfied the expression in question receives a deviant interpretation), and where movement can in principle take place prior to lexical insertion, it appears that nothing would in principle go wrong with movement into a theta position. Such movement can certainly still be banned by adopting additional assumptions (see in fact Chomsky 1995:312-316, who basically assumes that theta-relatedness is a base property), but the point is that additional assumptions are needed to ban it, not to allow it. There is also some empirical motivation that such movement should not be banned. Consider the following example from Chilean Spanish, noted by Gonzalez (1988) and discussed in this context by Bošković (1994).

(4)   A Juan le quiere gustar   Marta.
        to Juan cl. wants to-please Marta
        'Juan wants to like Marta.'

Juan bears an inherent case assigned by *gustar* and realized morphologically as *a*. Since inherent case is assigned together with a theta role (Chomsky 1986b), *Juan* must then have started in the embedded clause, where it got case and a theta role from *gustar* (*quiere* cannot assign the inherent case in question, cf. (*A) *Juan le quiere la fama* 'Juan wants fame', which rules out a control analysis for (4)). *Juan* is in fact interpreted as the experiencer of *gustar*. However, it is also interpreted as bearing the subject theta role of

quiere (and subjecthood tests show that it is located in the matrix SpecIP). Given this, Bošković (1994) argues that (4) should be analyzed in terms of movement into a theta position, where *Juan* moves from an embedded clause theta position (where it gets *a*) to a matrix theta position. While the possibility of movement into a theta position is still a controversial issue, it has been argued for by a number of authors and argued to have an impact on a number of important phenomena, including control, binding theory, and parasitic gaps. Thus, Hornstein (1999) analyzes all traditional control structures in terms of movement into a theta position, with the traditional controller starting in the place of PRO and moving into a higher theta position (see chapter 15).

The most recent theorizing has raised another issue regarding the basic design of language. Both GB and early Minimalism assume that there is a single point of spell-out, where the derivation separates into the phonology branch and the semantics branch. But why should that be the case? While under Chomsky's (1993) single spell-out approach PF and LF considerations determine when spell-out applies,[1] the assumption that spell-out applies only once still hides a trace of SS. A number of authors (see especially Uriagereka 1999, Epstein 1999, Epstein et al 1998, and Chomsky 2000, 2001a,b) explore the possibility that Spell-Out may apply multiple times. This leads to a radically derivational nature of the computation, where the interfaces access syntactic computations as the derivation proceeds without the mediation of PF and LF levels of representation. Syntactic derivation itself plays the central role in this approach, as there are no real levels of representation. The approach has a number of conceptual advantages. Instead of cyclic syntactic construction of a complete phrase-marker which is followed by a cyclic semantic composition that retraces syntactic steps, semantic composition can now proceed concurrently with syntactic structure building. Furthermore, MP assumes that syntactic computation is driven by the need to eliminate from the computation objects that are illegitimate at the interfaces, which interfaces cannot deal with. Given this, Epstein et al (1998) raise a question: how can a syntactic operation at some intermediate point of the derivation be driven by the notion of legitimacy at the interfaces, if these interfaces can access only the final PF and LF representations? On the other hand, in the derivational multiple spell-out model where the interfaces interface with the syntax throughout the syntactic derivation, the problem does not arise. In addition to the conceptual advantages, the dynamic multiple spell-out model also has a good amount of empirical justification, see e.g. Uriagereka (1999), Chomsky (2000), Franks and Bošković (2001), Simpson (2002), and Fox and Pesetsky (2005). As an illustration, consider the well-known problem regarding the blocking effect of wh-traces on *to*-contraction.[2]

(5)   a. *You wanna John go there.
        b. cf. You want John to go there.
        c. You wanna go there.
        d. cf. You want PRO to go there.
        e. *Who do you wanna go there?
        f. cf. Who$_i$ do you want t$_i$ to go there?

(5a-d) show that while a phonologically null subject of an infinitive does not block *to*-contraction, a lexically realized subject blocks it. A wh-trace apparently patterns with the latter in that it blocks *to*-contraction, as shown by (5e-f). If *to*-contraction is a PF phenomenon, (5e) is problematic: since wh-trace is phonologically null, we would expect it not to interfere with *to*-contraction, just like PRO does not interfere with it in (5c-d).[3] Under the multiple spell-out hypothesis, (5e) can be accounted for in a

---

[1] In particular, movements that take place prior to spell-out are assumed to be driven by the need to eliminate strong features, which are illegitimate PF objects, before syntactic information is sent to PF. The Principle Procrastinate, which favors LF movement, delays all movements that are not motivated by PF considerations until LF.

[2] The argument for multiple-spell out based on *to*-contraction can be traced back to Bresnan (1971).

[3] See Boeckx (2000) for discussion of the *to*-contraction paradigm under the movement approach to control.

principled way. Suppose that instead of waiting for the end of the syntactic derivation the phonology indeed interfaces with the syntax throughout the derivation. If what is sent to the phonology in (5e) is the matrix VP, *who* rather than a trace of *who* intervenes between *want* and *to* at the point when the phonology access the structure.

(6)   [$_{VP}$ You want who to go there]

It is then not surprising that (5e) patterns with (5a) rather than (5c) regarding *to*-contraction. An important question arises in this respect: which pieces of syntactic structure are sent to the interfaces? Chomsky (2000) argues that, in contrast to active VPs, passive VPs are not sent to the interfaces. This assumption enables us to account for the well-known fact that NP-traces do not block contraction, as shown by the possibility of contraction across the NP-trace in (7).

(7)   a. John is sposta leave.
        b. John$_i$ is supposed t$_i$ to leave.

If, in contrast to the matrix VP in (5), the matrix VP in (7) is not sent to the phonology, the phonology will only access (7) once the infinitival subject has moved from the embedded SpecIP, where it would intervene between *to* and the higher verb. As a result, the contraction is not blocked in (7) (the issue of which pieces of syntactic structure are sent to the interfaces under the multiple spell-out hypothesis is currently under debate, see section 2.3 for relevant discussion).

Another very influential proposal which has to do with the way syntax interacts with the interfaces concerns Kayne's (1994) proposal that linear order should be deduced from the hierarchical structure, where (roughly) if X asymmetrically c-commands Y, X precedes Y. While Kayne applies the algorithm in question in the syntax, Chomsky (1995) proposes that the algorithm should apply in PF, i.e. on the mapping from syntax to the PF interface. An interesting consequence of this proposal is that linear order is purged from the syntax (overt syntax and LF, as well as semantics). Syntax is then simply characterized by hierarchical structure, linear order being imposed on it only in PF, presumably as a requirement of the PF interface. This approach to linearization is an example of the minimalist agenda to reduce linguistic constraints to bare output conditions, which are determined by the external systems that the language faculty interfaces with.

Kayne's hypothesis that ties word order to structural relations has many consequences (see chapter 18). Thus, since the basic spec-head-complement structure can only be mapped to the order spec-head-complement, SOV languages like Japanese must be analyzed as involving movement of the object (without the movement, the verb would have to precede the object, the two being in a head-complement relation).

2.2. Last resort

Minimalism has insisted on the last resort nature of movement from its inception: In line with the leading idea of economy, movement must happen for a reason, in particular, a formal reason. Case may provide one such driving force. Consider (8).

(8)   Mary is certain *t* to leave

Mary cannot be case-licensed in the position of *t*. Raising to matrix SpecIP rectifies its case inadequacy, since the position into which raising has taken place licenses nominative. Once *Mary* has been case-licensed, it is no longer available for A-movement, to a case or a non-case position. This follows from Last Resort, if A-movement is driven by case considerations. Since *Mary* is case-licensed in the position of *t* in (9), Last Resort blocks further movement of *Mary*.

(9)   a. *Mary is certain *t* will leave
        b. *The belief Mary to be likely *t* will leave

One fruitful line of research regarding Last Resort concerns the issue of where the formal inadequacy driving movement lies, in the target (Attract) or the moving element (Greed). Greed was the earliest approach (Chomsky 1993), revived recently in Bošković (2007) based on considerations of successive cyclic movement. Under this approach X can move only if X has a formal inadequacy, and if the movement will help rectify the inadequacy. Under pure Attract, the target head always triggers movement (Chomsky 1995), which means the target must always have a formal inadequacy to be rectified by the movement (see also Lasnik 1995 for a combination of the two approaches to Last Resort). Under this approach, movement of *Mary* in (8) is driven by T/I: the head has a property, namely the EPP, that must be satisfied by an NP which triggers movement of *Mary* (*Mary*'s case-checking is merely a beneficial side effect of the satisfaction of the attractor's requirement.) While the Greed approach easily handles (9), such constructions require additional assumptions in a target-driven system. Chomsky (2000) posits the Activation Condition, which says X can move only if X has an uninterpretable feature, i.e. a formal inadequacy (the need to license its case feature in the case of A-movement). The approach is still sneaking in Greed into a system where movement is supposed to be target-driven. A target-driven approach, however, has a conceptual advantage in that under this approach the driving force for movement can be satisfied immediately (as soon as the relevant element enters the structure), rather than indefinitely later in the derivation, as in the moving-element driven approach. As discussed in Bošković (2011), there are, however, cases of movement that seem to be quite clearly driven by the properties of moving elements, like quantifier raising (which does not take place because the targets of quantifier raising would require an adjoined quantifier) or multiple movements to the same position such as multiple wh-fronting in languages like Romanian (see (19) below), where the very fact that all wh-phrases must front indicates that there is something wrong with wh-phrases that do not undergo movement.

2.3. Government

One early Minimalist objective was to re-examine all posited relations from GB, with the goal of eliminating those that were arbitrary. Attention immediately focused on government, a structural relation that was central in almost every module of GB. While this provided substantial unification it came at a significant cost, government being an arbitrary notion that did not follow from anything. There are many definitions of government in the literature. This in itself reveals its arbitrary nature that can be illustrated by the following definition of government, which is a modified version of Chomsky's (1986a) barriers system.

(10)   A governs B if there is no XP intervening between A and B, where XP is not a complement argument other than IP, and XP does not immediately dominate YP, where YP is not a complement argument.

Government was assumed to be crucially involved in Case Theory, providing a unification of the three configurations under which case was assumed to be assigned (spec-head for nominative case, head-complement for accusative case, and head-spec of the complement for exceptional accusative, as in *Mary believes him to be smart*), Binding Theory (the domain in which anaphors must be bound and pronouns must be free crucially had to contain a governor for the anaphor/pronoun; since PRO was assumed to be subject to binding conditions government was also crucially involved in the distribution of PRO), the distribution of null complementizers (which were assumed to be subject to a government requirement), and locality of movement (which is taken rather broadly here to include the licensing of traces). A great deal of effort went into providing accounts of these phenomena that do not appeal to government, in an attempt to eliminate the notion from the grammar. The research program has proved rather successful. Since movement has been argued to determine the distribution of anaphors and pronouns (see, e.g. Chomsky 1993, Hornstein 2001, and Kayne 2003), and the distribution of PRO has been argued to be determined either by locality of movement (under Hornstein's 1999 approach, where the controller of the traditional PRO undergoes movement) or Case theory (under the null case approach, where PRO bears a special case, see Chomsky and Lasnik 1993, Martin 1996 and Bošković 1997), the locality of case

licensing and the locality of movement have become the crucial cases (see also Bošković and Lasnik 2003, Pesetsky and Torrego 2001, and Landau 2007 for elimination of government from the domain of the distribution of null complementizers). In early Minimalism, all case licensing was reduced to the spec-head relation. What made that possible was adoption of object shift, an operation which moves accusative NPs to a higher position in which they enter a spec-head agreement relation with the verb. The operation was also assumed to apply to accusative subjects of infinitival complement clauses, which would then undergo A-movement motivated by case-licensing outside of their infinitival clauses. Some evidence for this is provided by examples like (11), where the subject of the infinitive binds an anaphor within a matrix adverbial, which indicates that it has moved into the matrix clause at the relevant point (see Postal 1974 and Lasnik and Saito 1991; there is some controversy regarding whether the movement is covert or overt, the latter possibility requiring overt short V-movement in English to derive the correct word order).

(11)   John proved [Mary and Jane to be innocent] during each other's trials.

Regarding locality of movement, early proposals attempted to deduce locality-of-movement effects from Chomsky and Lasnik's (1993) Minimize Chain Links Principle (MCLP), which requires each step of movement to be as short as possible.[4] Rizzi's (1990) relativized minimality effects then follow because such effects always involve skipping of a potential landing site, for example in *Mary seemed it was told t he left, Mary* has to skip embedded SpecIP when moving to the matrix SpecIP.[5]

More recently, Chomsky (2000, 2001a,b) has proposed another approach to locality which plays a crucial role in a number of phenomena, including locality of movement and case licensing (for relevant discussion, see also chapter 17). The approach is based on the notion of phase, the leading idea being that only the edge of a phase (i.e. spec and head of the relevant phrase) is available for operations outside of the phase (the Phase-Impenetrability Condition (PIC)). Assuming that CP and DP are phases, a wh-phrase moving out of a CP and DP then can only move outside of these phrases if it first moves to SpecCP and SpecDP.[6] If these positions are filled, making such movement impossible, as in (12), the PIC is violated.

(12)   a. ??What do you wonder [$_{CP}$ why John bought t]
          b. ?*What do you like [$_{DP}$ Mary's picture of t])

The PIC is also assumed to constrain the locality of case licensing. Thus, *John believed that him left* is ruled out because case-marking of the embedded clause subject by the matrix verb violates the PIC. In a multiple spell-out system, phases are also assumed to determine which pieces of structure are sent to the interfaces. They are also taken to determine the contents of a *numeration*. Chomsky (1995) argues that the computational system does not have direct access to the lexicon throughout the derivation; but only to a collection of lexical items, i.e. numeration, which serves as the starting point for the derivation. Not all lexical items in a sentence come from a single numeration. Rather, Chomsky argues that numeration corresponds to phases, with each phase built from a separate numeration. The concept of numeration is itself important since economy conditions are taken to compare only derivations with the same starting point, i.e. numeration. If this were not the case, silence would always be most economical since it does not require any syntactic operations. One non-trivial question concerns the proper definition of phases; the reader is referred to chapter 17 for this issue.

---

[4] There is a fine line to walk on here, given Bošković's (1994) proposal that movement cannot be too short, a line of research that Grohmann (2003), Abels (2003), and Boeckx (2007), among others, have expanded on (Grohmann calls the relevant constraint anti-locality).

[5] Takahashi (1994) deduces even the ban on extraction out of subjects and adjuncts from MCLP, the idea being that the MCLP forces the element moving out of a subject/adjunct to adjoin to the subject/adjunct, which leads to a violation of independent requirements.

[6] Rackowski and Richards (2005) and den Dikken (in press), however, question successive cyclic movement via SpecCP.

2.4. The Inclusiveness Condition

Chomsky (1993) proposes the Inclusiveness Condition, which confines the power of syntax to (re)arrangements of lexical items, banning syntax from creating new objects. The condition is very appealing conceptually due to its restrictive nature. It has also led to re-examinations of several phenomena that have yielded very important results. I will discuss here its effects for phrase structure and the trace theory of movement, starting with the latter.

2.4.1 The copy theory of movement

It is a fundamental property of human language that syntactic elements can be interpreted in positions different from the ones where they are pronounced. In GB such dependencies were implemented by using movement and traces. Thus, *Mary*, which is interpreted as object of *kiss* in (13), is generated in this position and then it moves, leaving being a co-indexed trace, a phonetically null element that has the interpretational properties of the moved element.

(13) $Mary_i$ was kissed t_i

Traces are prime examples of creationism in syntax and, as such, violate the Inclusiveness Condition. Chomsky (1993) therefore abandons trace theory and resurrectss the copy theory of movement, according to which the trace of an overtly moved element is a copy of the moved element that is deleted in PF but available for interpretation in LF. (Copies are annotated here with superscripts.)

(14) $Mary^k$ was arrested ~~$Mary$~~^{k}.

Chomsky (1993) shows that in addition to conforming to the Inclusiveness Condition, the copy theory considerably simplifies the analysis of reconstruction phenomena (see below). Furthermore, by making it possible to treat reconstruction as an LF phenomenon, the copy theory contributes to the research attempt to eliminate non-interface levels of representation. Another attractive feature of the copy theory is that, by eliminating traces, it reduces the number of theoretical primitives.

However, the copy theory has brought in a new set of questions: it now has to be determined which copies reach the interfaces and why this is so. The positions in which elements undergoing movement are pronounced and interpreted are generally unique, and not necessarily identical. To ensure this under the copy theory, it is standardly assumed that only one copy of X remains at the interface levels. The question is then which copy should survive deletion.

On the LF side, there is at least some choice in deciding where deletion should take place in nontrivial chains, with a preference for deletion of the restriction of the wh-operator in the head of operator-variable chains. Consider (15). Chomsky (1993) argues that the upstairs reading of *himself* is obtained after the tail of the *wh*-chain is deleted (see (16a), where *himself* is locally bound by *Joe;* note that if theta-roles are features that can be checked derivationally the deletion may not need to leave anything behind.) Under the downstairs reading, *himself* is deleted in the head of the *wh*-chain and remains in its tail (see (16b), where *himself* is locally bound by *Jim*.)

(15) Joe_i wondered which picture of $himself_{i/j}$ Jim_j bought
(16)   a. Joe wondered [$_{CP}$ [which picture of himself]^k [$_{IP}$ Jim bought ~~[which picture of himself]~~^{k}]]
        b. Joe wondered [$_{CP}$ [which ~~picture of himself~~]^k [$_{IP}$ Jim bought [~~which~~ picture of himself]^k]]

Chomsky argues that there is a preference for minimizing operator restriction in LF, which normally leads to deletion in the head of Ā-chains. This is motivated by the impossibility of coreference between *he* and *Tom* in (17). To exclude (17) while allowing the upstairs reading of *himself* in (15), Chomsky suggests that in (16a), *himself* undergoes LF anaphor movement into the matrix clause from the head of the *wh*-chain (see chapter x regarding anaphor movement); deletion in the head of the *wh*-chain is then blocked because it would break the anaphor movement chain. By contrast, in (16b) the lower copy of *himself*

undergoes LF anaphor movement within the lower clause so that the deletion within the head of the *wh*-chain is permitted.

(17)   a. *Mary wondered which picture of Tom$_i$ he$_i$ liked.
         b. *Mary wondered [$_{CP}$ [which ~~picture of Tom~~]^k he liked [~~which~~ picture of Tom]^k]

While on the LF side there is clearly some choice in deciding where deletion should take place in nontrivial chains, early on it was assumed that no such choice is available in PF, the head of a nontrivial chain always being the sole survivor, based on paradigms like (18).

(18)   a. [[the student]^k was arrested ~~[the student]~~^{k}]
         b. *[[~~the student~~]^k was arrested [the student]^k]
         c. *[[the student]^k was arrested [the student]^k]
         d. *[[the ~~student~~]^k was arrested [~~the~~ student]^k]
         e. *[[~~the~~ student]^k was arrested [the ~~student~~]^k]

Such paradigms led to adopting the stipulation that lower copies cannot be phonetically realized, which was rather unfortunate since it tacitly resuscitated traces. More recent work has, however, amassed considerable evidence that "traces", i.e. lower copies, can be phonetically realized. It turns out that just as in LF there is a preference for deletion in the head position of non-trivial chains (at least with operator-variable chains), deletion of lower copies in PF is just a preference, not the only option. In particular, Franks (1998) argues that a chain is pronounced in the head position, with lower copies deleted in PF, unless pronunciation in the head position would lead to a PF violation. If and only if the violation can be avoided by pronouncing a lower copy of the chain, the lower copy is pronounced and the head of the chain is deleted. There is a good amount of empirical evidence for this state of affairs, which has also been shown to be deducible from independent assumptions by Nunes (2004). One such argument is given in Bošković (2002a) based on multiple wh-fronting in Romanian (for an overview of other arguments and additional references, see Bošković and Nunes 2007; as shown there, there are in fact cases where different pieces of different chain links are realized, just as in the LF procedure discussed regarding (16b)).
     Romanian requires all *wh*-phrases to front in questions (19). However, the second *wh*-phrase does not appear to move if it is homophonous with the first fronted *wh*-phrase (20).

(19)   a. Cine ce   precede?
           who what precedes
         b. *Cine precede   ce?
            who precedes what
            'Who precedes what?'
(20)   a. Ce   precede ce?
           what precedes what
         b. *Ce   ce   precede?
            what what precedes
            'What precedes what?'

Many languages have low-level PF constraints against consecutive homophonous phrases, which is what rules out (20b). (20a) then represents an intricate interplay between phonology and syntax, with phonology apparently overriding syntax, more precisely, the need to satisfy a PF requirement appears to override the need to satisfy a syntactic requirement. The mechanism of pronunciation of lower copies provides a straightforward way of resolving this phonology-syntax conflict. Given that Romanian has a syntactic requirement that forces all *wh*-phrases to move overtly (Bošković 2002a argues this involves focalization), the second *wh*-phrase must move in the syntax, as in (21) (irrelevant copies are ignored).

(21)   [ce ce^i precede ce^i]

If the highest copy of the second *wh*-chain is pronounced, a PF violation obtains due to a sequence of homophonous *wh*-elemens. This is precisely the situation where a lower copy can be pronounced under Franks's (1998) approach to the pronunciation of non-trivial chains.

(22)   [ce ~~ce^i~~ precede ce^i]

By deriving (20a) from (21), the pronunciation of a lower copy analysis accounts for the contrast between (19b) and (20a) without violating the syntactic requirement that forces all *wh*-phrases to move overtly in Romanian and without look-ahead from the syntax to the phonology. It also resolves the problem of the phonology-syntax interaction raised by (20a), without having phonology override syntax.

There is also independent evidence that the second *ce* in (20a) indeed moves overtly. *In-situ wh*-phrases in multiple questions in languages like English differ from their moved counterparts in being unable to license parasitic gaps.

(23)   a. What did John read without filing?
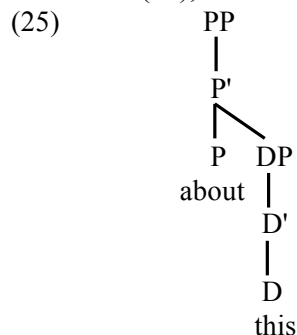   b. *Who read what without filing?

Significantly, the *in-situ wh*-phrase in (20a) does license parasitic gaps. The fact that the "*in-situ*" wh-phrase in (24) patterns with the moved wh-phrase in (23a) rather than the *in-situ* wh-phrase in (23b) is exactly what is expected if (20a) involves overt *wh*-movement and lower copy pronunciation.

(24)   Ce   precede   ce   fără   să   influenţeze?
         what precedes what without subj.particle influence.3p.sg
   'What precedes what without influencing?'

In addition to the issue of copy interpretation at the interfaces, the copy theory also has more general consequences for the operation Move. The basic structure building operation is Merge, which combines two things making one of them the head of the resulting structure (see section 2.4.2.) What Move does is take an element that is already merged into the structure, and copies it by merging it into another position. From this perspective, Move is not a primitive operation, but the combination of the operations Merge and Copy. While remerger of an element that is already merged into the structure is referred to as Internal Merge and initial merger of an element into the syntactic tree is referred to as External Merge there is then still only one basic structure-building operation, namely Merge. The mechanics of the operation will be discussed in detail in the next section.

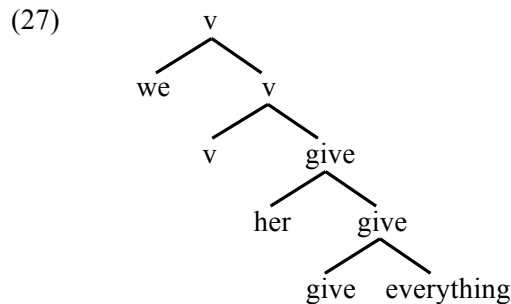### 2.4.2. Bare Phrase Structure and late insertion

Above, we have seen that the Inclusiveness Condition has led to the introduction of the copy theory of movement. The condition also has important consequences for phrase structure. Bar-levels, vacuous projections, even the terminal/non-terminal distinction, standard tools of the GB framework present in the GB tree in (25), all violate the ban on creationism in syntax (none of them are present in the lexicon).

(25)

```
            PP
            |
            P'
           ╱ ╲
          P   DP
        about  |
               D'
               |
               D
              this
```

This has led to a re-examination of phrase structure (and X'-theory), the result of the re-examination being known as Bare Phrase Structure (Chomsky 1995: chapter 4). Consider (25) from this perspective. In (25), two lexical items, *about* and *this,* are combined by the operation Merge, which takes two syntactic objects and forms a new syntactic object out of them. The resulting object is labelled by *about*, which means that the object is a projection of *about* (i.e. it is of the same type as *about*). This gives us the following representation for (25).

(26)   about
            /   \
       about   this
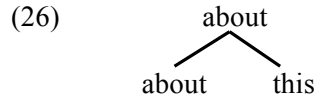
While (26) conforms with the Inclusiveness Condition, it raises a number of interesting questions regarding how the traditional distinctions between maximal and minimal projections, complements and specifiers can be implemented in this system, given that they cannot be identified by any special marking. Consider the issue with respect to a slightly more complicated case of a double object construction, where v introduces the external argument, with *give* raising to v (the movement is ignored here).

(27)   v
            /   \
       we   v
                  /   \
                 v   give
                        /   \
                      her   give
                                 /   \
                              give   everything

Following Muysken (1982), Chomsky argues for a relational definition of the relevant notions: a category that does not project any further is a maximal projection (e.g. the highest *give*), and one that is not a projection at all is a minimal projection (the traditional head, which is the lowest *give*). As for the complement/specifier distinction, Chomsky suggests that the most local relation with a head Y corresponds to the head-complement relation, all other relations (putting aside adjunction for the moment) being head-specifier. This means that the first merger with a head yields a complement (*everything* in (27)), all additional mergers involve specifiers (*her* in (27)).

Bare phrase structure has a number of very interesting consequences. First, it allows for the existence of elements that are at the same time heads and phrases. Chomsky suggests that clitics, which do seem to share some properties of both heads and phrases, are such elements (see here Bošković 2002b; note that some of the lexical items in (27) may actually have more internal structure than indicated in (27)). The system also naturally allows for the existence of multiple specifiers.[7] One of the instances where this possibility has been argued to be instantiated is multiple wh-fronting, as in Romanian (19a)), where under Koizumi's (1994) analysis each wh-phrase occupies a separate SpecCP.

This system leaves adjuncts in a somewhat uncomfortable position. One of the reasons behind the postulation of adjunction was to allow for the possibility of merging multiple phrases (more precisely, more than two phrases) within the same phrase. In the Bare Phrase Structure system this can be achieved by employing multiple specifiers. There are, however, other reasons to postulate adjunction, the most important of which may be the special status of adjuncts with respect to the cycle. (Adjuncts also

---

[7]More precisely, multiplicity of specifiers cannot be blocked in the Bare Phrase Structure system. In this respect, the system contrasts with Kayne (1994), which disallows multiple specifiers (see chapter 18).

typically fail to enter into agreement relations and have different case requirements from arguments.)[8] As discussed above, in MP the derivation proceeds cyclically bottom up starting with what ends up being the most deeply embedded structural unit. The requirement of strict cyclicity ensures, for instance, that it is not possible to get around the wh-island effect in (28a) by doing long distance wh-movement first (28b) and then filling the embedded clause SpecCP ((28c); for various approaches to cyclicity, see Lasnik 2006).

(28)   a. ??What do you wonder why John bought
        b. What do you wonder John bought why
        c. What do you wonder why John bought

Lebeaux (1988) argues that certain elements, in fact precisely those that have been assumed to enter the structure by adjunction, should be exempt from the cyclicity requirement. His argument is based on the following contrast:

(29)   a. *Which argument that John$_i$ is a genius did he$_i$ believe?
        b. Which argument that John$_i$ made did he$_i$ believe?

(29a), which involves a nominal complement, can be captured if the restriction of the operator must be reconstructed to the base position, i.e. the position it occupied prior to wh-movement (see in this respect the discussion of the preference for minimizing operator restriction regarding (17) above). What about (29b) then? The relevant element in (29b) is a relative clause, which is traditionally treated as an adjunct. Lebeaux observes that the grammaticality of (29b) can be accounted for if adjuncts can be inserted non-cyclically. The issue of reconstruction does not arise on the derivation on which the relative clause is introduced into the structure acyclically after wh-movement, since this derivation has only one copy of the adjunct, the one that is pronounced in the matrix SpecCP. Furthermore, Stepanov (2001) argues that adjuncts not only can be, but in fact must be inserted into the structure acyclically based on (30). In particular, the fact that, in contrast to (30b), condition A is apparently violated in (30a) provides evidence that the adjunct clause must in fact be adjoined after wh-movement (given late adjunction, *the lawyers* cannot bind *each other* at any point of the derivation).[9]

(30)   a. *What evidence that each other$_i$'s friends brought up in court did the lawyers$_i$ refuse to talk about?
        b. The lawyers$_i$ refused to talk about the evidence that each other$_i$'s friends brought up in court

---

[8] Regarding formal implementation of adjunction, adjuncts are assumed not to change the label and bar-level of their targets (see Chomsky 1995: chapter 4). The constituent represented by the tree in (i) is a non-minimal, maximal projection.

(i)
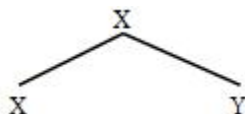
The addition of a specifier turns the constituent into a non-minimal, non-maximal projection. The addition of an adjunct, on the other hand, does not change the bar level of the target, which remains a non-minimal, maximal projection.

[9] There is also a line of research that accounts for special properties of adjuncts by assuming that adjuncts are not fully integrated into the structure (or they are on a separate plane) at the point when the processes in which adjuncts do not participate occur (see in this respect Chomsky 2001b).

Obligatory late insertion of adjuncts suggest an interesting way of capturing the ban on extraction out of adjuncts. Thus, if the adjunct in (31) must be introduced into the structure late, after wh-movement takes place, it follows that no wh-movement can take place out of the adjunct (the account raises a number of questions, see in this respect the discussion in Stepanov 2001; for alternative accounts of the Adjunct Condition, see Chomsky (1986b), Takahashi (1994), Nunes and Uriagereka 2000, Starke (2001), Boeckx (2003), and Müller (2010), among others; extraction from adjuncts is also discussed in chapter 17).

(31)   ?*What did John fall asleep [after Mary had bought]

    There is another interesting question concerning late insertion: can lexical insertion take place in PF and LF? In the GB framework it was tacitly assumed that all lexical insertion has to take place prior to S-Structure (i.e. prior to spell-out); otherwise, the sound-meaning relation would completely collapse. Thus, if free insertion in LF were allowed (32a) could mean (32b), with *John does not think* inserted in LF (hence it would be present in the semantic representation but would not be pronounced). On the other hand, if free insertion in PF were allowed (32b) could mean (32a), with *John does not think* inserted in PF (hence it would be pronounced but it would not have semantic representation).

(32)   a. Mary left
         b. John does not think that Mary left

Chomsky (1995) suggests a way of deducing the ban on PF/LF insertion. Consider the lexical item *John*. Like any lexical item, *John* is a set of features, with at least three kinds of features: formal (e.g. case feature), phonological (the instruction how to pronounce *John*), and semantic (the instruction how to interpret *John*). Given this, the ban on late (i.e. PF/LF) insertion follows from the Principle of Full Interpretation, which requires that whatever is present at the interface levels be interpretable by the interfaces. Thus, if *John* is inserted in LF the derivation crashes because LF cannot interpret the phonological features of *John*. If, on the other hand, *John* is inserted in PF, PF will not know how to interpret the semantic features of *John*. The only way to derive a legitimate PF and a legitimate LF is for *John* to be inserted before the "SS" is reached. PF will then strip off the phonological features of *John,* and the semantic features of *John* will proceed into LF. This line of reasoning allows lexical insertion to take place in PF and LF under certain conditions. To be more precise, it allows PF insertion of semantically null elements and LF insertion of phonologically null elements (as long as such insertion obeys the cycle, assuming the cycle holds in LF). The former may be instantiated in the case of semantically dummy elements like *do* of *do*-support. Bošković (1998) argues that the latter option is the best way to handle French wh-in-situ, the element that is inserted in LF being the interrogative C. As illustrated in (33)-(35), wh-in-situ is in fact possible in French only in the contexts where LF insertion is in principle possible, namely, when the C is phonologically null (compare (33a) and (35b)) and when it is located at the top of the tree (compare (33a) and (34b)). (Note that not all dialects of French allow overt C examples like (35a).)

(33)   a. Tu   as   vu   qui?
            you have seen whom
            'Who did you see?'
         b. Qui   as-tu   vu?
(34)   a. Pierre a demandé qui   tu   as   vu.
            Pierre has asked   whom you have seen
         b. *Pierre a demandé tu as vu qui.
(35)   a. Qui   que tu as   vu?
            whom   C   you have seen
            'Who did you see?'

b. *Que tu as vu qui?

2.5. Economy and optionality

The above paradigm can also be used to illustrate some issues regarding the minimalist guideline that language is economical. The French data in (33) bring up the question of optionality. Truly optional movement does not fit well with a framework like Minimalism where all movement takes place for a reason: in such a framework either there is a reason for a movement to take place, in which case it must take place, or there isn't, in which case it in fact cannot take place, given Last Resort. Everything else being equal we then would not expect to find truly optional movements. In cases like (33), it then must be the case that everything else is not equal. Under the above analysis this is indeed the case. Notice first that we cannot assume that French interrogative C optionally triggers movement. If that were the case, we would not be able to ever enforce the +wh-movement option, which would leave the ungrammaticality of (34b) and (35b) unaccounted for. Bošković (1998) then argues that French interrogative C in fact requires wh-movement. The difference between (33a) and (33b) is in the timing of C-insertion. In both cases, the C triggers wh-movement immediately upon its insertion (note in this respect that French wh-in-situ shows locality-of-movement effects): however, the C is inserted overtly in (33b), hence it triggers overt wh-movement, and it is inserted covertly in (33b), hence it triggers covert wh-movement (in other words, there is no overt wh-movement in (33a) because its trigger is not present in the structure at the relevant point of the derivation). In (34) and (35), where the LF C-insertion derivation is blocked (it is blocked in (35b) because the C is phonologically overt and in (34b) because the insertion would not take place at the top of the tree, violating the cycle), overt C insertion and overt wh-movement are the only options. The above discussion illustrates what one would expect to find in the cases involving seemingly optional movements: not everything is equal in the movement and non-movement options.

The French data in question also show that economy should be evaluated locally, without look-ahead. Chomsky (1993) argues that covert movement is more economical than overt movement, hence should be preferred to overt movement when both are in principle possible. From this perspective, the LF C-insertion derivation should be more economical than the overt C insertion derivation, since it triggers covert wh-movement, which is more economical than overt wh-movement. It then appears that (33a), the LF C-insertion derivation, should block (33b), the overt C-insertion derivation. Notice, however, that this is the case only if look-ahead is allowed. C-insertion itself works in exactly the same way whether it is covert or overt; there is no reason to prefer one of these options. It just happens that later down the road one of these operations will lead to a more economical movement operation. If economy considerations must be evaluated locally, without look-ahead, rather than globally, with look-ahead, we then get exactly the right outcome, with both (33a) and (33b) allowed. The data in question thus favor local economy over global economy, as one would expect from a system with efficient design (see Collins 1997 for much relevant discussion).

The system then does not always choose one option. As noted by Chomsky (1991), more than one option is allowed if they are equally economical. In this respect, Superiority effects provide a rather nice confirmation of the idea of economy in language. Consider (36), whose structure prior to wh-movement is given in (37).

(36)   a. Who did you tell that she should buy what?
         b. ?*What did you tell who that she should buy?
(37)   You tell who that she should buy what

The GB account of the contrast in (36) was based on Chomsky's (1973) Superiority Condition (see (38)), which basically says that if X and Y are candidates for a particular movement operation, and X asymmetrically c-commands Y, then X rather than Y undergoes the movement in question (see chapter 17; there are actually other pre-minimalist accounts of (37); however, they all make the same prediction as (38) with respect to the Bulgarian data discussed below.)

(38)   No rule can involve X, Y in the structure ..X...[...Z...WYV...] where the rule applies ambiguously
          to Z and Y, and Z is superior to Y. The category A is superior to the category B if every major
          category dominating A dominates B as well but not conversely.

Turning to minimalism, the Superiority effect in (36) is perhaps the most straightforward illustration of
economy of derivation, which requires that every requirement be satisfied through the shortest movement
possible. English interrogative C requires its Spec position to be filled by a wh-phrase, however this
requirement is to be implemented. In (37) the requirement can be satisfied by moving either *who* or *what*.
Since the former results in shorter movement, (36a) is preferred to (36b).[10]
          While the economy account is simpler hence conceptually preferable to the account based on (38)
there is also an empirical reason to prefer the economy account. Consider the Bulgarian MWF data in (39).

(39)   a. *Koj e vidjal kogo?
               who is seen   whom
          b. Koj   kogo e vidjal?
             'Who saw whom?'
(40)   a. Kogo kakvo e pital   Ivan?
             whom what   is asked Ivan
             'Whom did Ivan ask what?'
          b. ?*Kakvo kogo e pital Ivan?

In Bulgarian MWF all wh-phrases move to SpecCP. The wh-phrase that is highest prior to wh-movement
(the subject in (39) and the indirect object in (40)) must come first in the linear order. This has been
analyzed as a Superiority effect, which means that the highest wh-phrase moves first, the second wh-
phrase either right-adjoins to the first wh-phrase, as in Rudin (1988), or moves to a lower SpecCP, as in
Richards (2001). As discussed in Bošković (1999), like English +wh C, Bulgarian +wh C requires
movement to its Spec. In addition, there is a focusing requirement on wh-phrases, which is also satisfied
by moving wh-phrases to SpecCP. Consider now what should happen in constructions involving three
wh-phrases. Let us assume that whichever wh-phrase moves first to SpecCP satisfies the requirement of
the interrogative C to have a filled SpecCP. As a result, the wh-phrase that is highest prior to wh-
movement must move first to SpecCP; this way the formal inadequacy in question is satisfied in the most
economical way. The lower two wh-phrases then move to SpecCP to satisfy their focusing requirement.
Whichever order these two wh-phrases move in, the requirements are satisfied through the movements of
the same length, i.e. the movements always cross the same maximal projections. Under the economy
account we would then expect that the highest wh-phrase will move first, and then the other two wh-
phrases can move in any order. On the other hand, the account based on (38) will always force the higher
wh-phrase to move first, which means that the order of all three fronted wh-phrases should be fixed,
reflecting their order prior to wh-movement. As discussed in Bošković (1999), the prediction of the
economy account is borne out: the highest wh-phrase must move first, the order of the second and the
third wh-phrase is free. Thus, while the dative must move before the accusative in (40), this is not the case
in (41), where their order of movement is free. (Any order where the subject is not first is unacceptable.)

(41)   a. Koj kogo   kakvo e pital?
             who whom what   is asked
             'Who asked whom what?'
          b. Koj kakvo kogo e pital?

---

[10]The account can actually be extended to all relativized minimality effects. Thus, the problem with (28a) is that
what undergoes wh-movement to the matrix SpecCP although there is a wh-phrase that is closer to the matrix
SpecCP than *what*, namely *why*.

The above data provide an argument for the economy account of Superiority and for the general notion of economy in language. It also shows that economy does not always rule out all but one option. In cases where several options are equally economical they are all allowed.

2.6. Covert movement/dependencies

A lively area of research in current minimalist theorizing involves the nature of covert movement, or more generally, covert dependencies. In GB there was a standard treatment of covert movement. Covert movement was pretty much the same as overt movement (putting aside the timing difference, the former taking place after S-Structure and the latter before S-Structure), except that it was assumed not to be subject to some constraints, like Subjacency (see Huang 1982; see also chapter 22 for alternative views). Thus, Japanese (42) was assumed to involve movement of the wh-phrase to SpecCP, just like its English counterpart, the only difference being that the movement was assumed to take place in LF in (42).

(42)   Anata-ga dare-o   mita ka
        you-nom who-acc saw Q
         'Who did you see?'

Minimalism initially kept this treatment of covert movement. However, Chomsky (1994) suggested that covert movement involves movement only of formal features. The reason behind this suggestion was the general economy principle that movement should carry as little material as possible; since movement is assumed to be motivated by the need to license/check formal features only formal features should then be affected by movement. Chomsky (1994) suggested that this is indeed what happens in cases like Japanese (42). However, he argued that pure formal feature movement is not a possibility prior to SS. The reason for this is that such movement feeds pronunciation, and lexical items whose features are scattered in different positions (with formal features separated from other features due to feature movement) are unpronounceable.[11] Chomsky (2000), however, reanalyzes covert dependencies in terms of Agree, which is a feature checking operation that takes place at a distance without actual movement. Under this approach the interrogative C and the wh-phrase in (42) simply undergo feature checking via Agree without actual movement.[12] There are two additional treatments of covert dependencies, both of which arose as a result of elimination of the traditional LF component and the LF cycle. As noted in section 1, the most recent approach to the basic design of grammar eliminates LF. Covert and overt movement then cannot be distinguished by sequential ordering in the derivation: since there is only one cycle, they both have to take place on the same cycle.[13] One way to distinguish covert and overt movement is then in terms of which copy is pronounced (see section 2.4.1): the highest would be pronounced in the latter and the lowest in the former (see Bobaljik 1995 and Groat and O'Neil 1996). Another way is to assume that covert and overt movement differ with respect to the timing of the transfer to the PF interface (see Nissenbaum 2000): they both take place on the same cycle, but the phrase hosting the movement can be transferred to the PF interface either after (in the case of overt movement) or before (in the case of covert movement) the relevant movement takes place. Notice finally that the Agree approach is also consistent with the one-cycle syntax model.

Determining the proper treatment of covert dependencies is currently a lively area of research. Empirical facts however suggest that there may not be a single way of treating such dependencies. Thus,

---

[11]It is actually not clear that movement of formal features prior to SS would cause a pronunciation problem since phonological features (and this is what PF cares about) would still be present together.

[12]Another non-movement approach is to assume that the Q unselectively binds the wh-phrase. This treatment, however, is not an option for all covert dependencies, which other options mentioned in this section at least in principle can be.

[13]What was discussed in section 1 in this respect was the multiple-spell out model. Adoption of the single cycle model actually does not require adoption of multiple spell-out; see for example Bobaljik (1995), Groat and O'Neil (1996).

wh-in-situ does not exhibit uniform behavior crosslinguistically (compare, e.g. the behavior of the wh-in-situ in (23b) and (24) with respect to parasitic gap licensing), which may be difficult (though not necessarily impossible) to account for if all covert dependencies are treated in the same way.

2.7. Features and feature checking

Let us now look more closely at feature checking/licensing, which drives much of the syntactic computation in minimalism. As discussed above, lexical items enter the computation as sets of features. Some of these features require feature checking, i.e. establishing relations with other features, during the syntactic computation.

There are two properties of features that are important in this respect. Chomsky (1995) introduces the interpretable/uninterpretable feature distinction, the criterion here being semantic interpretability, or more precisely, interpretability at the LF interface. Case, for example, is taken to be an uninterpretable feature. Other features are more complicated in that they are uninterpretable only on some lexical items. For example, the number feature is uninterpretable on the verb, but interpretable on the noun. (Thus, the subject in *A student likes Mary/Students like Mary* is interpreted differently depending on whether it is singular or plural, which is not the case with the verb.[14]) Given the Principle of Full Interpretation, which requires that everything that is present at the interface levels has to be interpretable by the interfaces, uninterpretable features then have to be eliminated from the structure before the LF interface. This is accomplished through feature checking, an operation which will be discussed below.

Chomsky (2000) introduces another important distinction: valued vs unvalued features. Consider (43)-(45) (*kola* is a plurale tantum).

(43) Zelena   kola   su kupljena.
      green.fem car.fem are bought.fem
      'The green car was bought.'

(44) Zeleno   auto   je kupljeno.
      green.neut car.neut is bought.neut

(45) Zeleni   automobil je kupljen.
      green.masc car.masc   is bought.masc   (Serbo-Croatian)

The gender of the adjective and the participle depends on the gender of the noun. *Green* can be feminine, neuter, or masculine; which gender it has depends on the noun it modifies. On the other hand, nouns have a fixed gender specification: *kola* is always feminine, *auto* neuter, and *automobil* masculine. As noted by Pesetsky and Torrego (2008), this can be captured if adjectives and participles are lexically unvalued for gender: they receive their gender value after undergoing agreement with a noun that already has a valued gender. Since the gender of the noun is lexically valued it does not depend on the noun's syntactic context. Under the plausible assumption that unvalued features are illegitimate at the interfaces (the interface has to know the value of a feature to be able to interpret it), only valued interpretable features are legitimate at the LF interface. Other features need to be eliminated before reaching the interface. In principle, that could be accomplished through deletion of relevant features. However, Chomsky (2001a) posits that only valued features can be deleted. Unvalued uninterpretable features then need to be valued before they can be eliminated. From this perspective, what drives the feature checking operation is valuation, i.e. the need to value unvalued features.

An element X with an unvalued feature K then probes its c-command domain to find Y with a valued K. Once X finds Y, Y values K of X. The process in question is the Agree operation. If X's K is uninterpretable, which means the LF interface cannot deal with it, it will then be deleted; if it is interpretable, it will remain in the structure. A sentence like (46) then has the following derivation:

(46)   John left

---

[14] The exception here is the number of pluralia tantum nouns like *kola* in (43), which does not have semantic interpretation.

T	John
uninterpretable unvalued phi features	valued interpretable phi features
	unvalued uninterpretable Case feature

T in (46) triggers the Agree operation. The valued phi features of *John* value the phi features of T, after which the phi features of T are deleted. Only phi features of *John*, which are interpretable at the LF interface, then reach the LF interface level. Notice also that *John* has an unvalued Case feature. Chomsky (2001a) assumes that T does not have a case feature, the case feature of *John* being valued as nominative as a reflex of the phi-feature checking relation with finite T. It is also possible to implement case-licensing in (46) without the stipulatory notion of reflex checking. Thus, as noted by Bošković (in press), it can be assumed that T has a valued case feature (nominative; after all finite T always licenses nominative, while the case feature of a noun depends on its syntactic context[15]), with the case feature of T valuing the case feature of *John*. Since deletion of valued uninterpretable features is allowed, both the case feature of *John* and the case feature of T can be deleted following the Agree relation so that they do not reach the LF interface. Finally, the Agree relation in (46) is followed by movement of *John* to SpecTP, which is triggered by a property of the target in the target-driven system (the EPP property), or by a property of the moving element in the moving-element driven system (for example the requirement that unvalued uninterpretable features function as probes, which means that they must c-command their valuators, see Epstein and Seely 2006 and Bošković 2007).

Recall also that valued interpretable features of *John* are not deleted after undergoing a feature-checking Agree relation. This implies that they should be able to undergo such a relation more than once, an option which is realized in multiple verb constructions where each verb agrees with the subject.

(47)   I   ragazzi sono   stati   invitati.
         the boys   are.pl been.pl invited.pl
         'The boys were invited.'   (Italian)

3. Conclusion
The leading idea of the Principles and Parameters approach is the existence of Universal Grammar, a genetic endowment which helps children acquire language that comprises a set of principles, which are invariant across languages, and a set of parameters, which provide limited options for language variation. The Government and Binding Theory and the Minimalist Program are two implementations of the Principles and Parameters approach, with the Minimalist Program building upon the Government and Binding Theory. Minimalism attempts not only to explain the properties of the language faculty but also to answer the question why these properties are the way they are.

References

Abels, Klaus. 2003. Successive cyclicity, anti-locality, and adposition stranding. Doctoral dissertation, University of Connecticut, Storrs.
Bobaljik, Jonathan. 1995. Morphosyntax: The syntax of verbal inflection. Doctoral dissertation, MIT, Cambridge, Mass.
Boeckx, Cédric. 2000. A note on contraction. *Linguistic Inquiry* 31: 357-366.
Boeckx, Cédric. 2003. *Islands and chains: Resumption as stranding*. Amsterdam: John Benjamins.
Boeckx, Cédric. 2007. *Understanding minimalist syntax*. Oxford: Blackwell.
Borer, Hagit. 1984. *Parametric syntax*. Dordrecht: Foris.
Bošković, Željko. 1994. D-structure, theta-criterion, and movement into theta-positions. *Linguistic*

---

[15] Chomsky disallows the possibility of valued uninterpretable features essentially by stipulation. The possibility, however, needs to be allowed on empirical grounds (see Pesetsky and Torrego 2007 and Bošković in press). Thus the gender feature of the nouns in (43)-(45) is valued, as discussed above, but it is also semantically uninterpretable, gender in Serbo-Croatian being an arbitrary, grammatical feature. As discussed in the text, allowing valued uninterpretable features also makes possible elimination of reflex feature checking.

Analysis 24: 247-286.

Bošković, Željko. 1997. *The syntax of nonfinite complementation: An economy approach*. Cambridge, Mass.: MIT Press.

Bošković, Željko. 1998. LF movement and the Minimalist Program. In *NELS* 28, ed. by Pius N. Tamanji and Kiyomi Kusumoto, 43-57. GLSA, University of Massachusetts, Amherst.

Bošković, Željko. 1999. On multiple feature-checking: Multiple wh-fronting and multiple head-movement. In *Working Minimalism*, ed. by Samuel Epstein and Norbert Hornstein, 159-187. Cambridge, Mass: MIT Press.

Bošković, Željko. 2002a. On multiple wh-fronting. *Linguistic Inquiry* 33: 351-383.

Bošković, Željko. 2002b. Clitics as non-branching elements and the Linear Correspondence Axiom. *Linguistic Inquiry* 33: 329-340.

Bošković, Željko. 2007. On the locality and motivation of Move and Agree: An even more minimal theory. *Linguistic Inquiry* 38: 589-644.

Bošković, Željko. 2011. Last resort with Move and Agree in derivations and representations. In *The handbook of linguistic minimalism*, ed. by Cédric Boeckx. Oxford University Press.

Bošković, Željko. In press b. On valued uninterpretable features. *Proceedings of NELS 39*.

Bošković, Željko, and Howard Lasnik. 2003. On the distribution of null complementizers. *Linguistic Inquiry* 34: 527-546.

Bošković, Željko, and Jairo Nunes. 2007. The copy theory of movement: A view from PF. In *The copy theory of movement*, ed. by Norbert Corver and Jairo Nunes, 13-74. Amsterdam/Philadelphia: John Benjamins.

Bresnan, Joan W. 1971. *Contraction and the transformational cycle.* Ms., MIT, Cambridge, Mass.

Chomsky, Noam. 1973. Conditions on transformations. In *A festschrift for Morris Halle*, ed. by Stephen Anderson and Paul Kiparsky, 232-286. New York: Holt, Rinehart and Winston.

Chomsky, Noam. 1986a. *Barriers*. Cambridge, Mass.: MIT Press.

Chomsky, Noam. 1986b. *Knowledge of language*: *Its nature, origin, and use*. New York: Praeger.

Chomsky, Noam. 1993. A minimalist program for linguistic theory. In *The view from Building 20*, ed. by Kenneth Hale and Samuel J. Keyser, 1-52. Cambridge, Mass.: MIT Press.

Chomsky, Noam. 1994. Bare phrase structure. In *MIT Occasional Papers in Linguistics* 5. MITWPL, Department of Linguistics and Philosophy, MIT, Cambridge, Mass.

Chomsky, Noam. 1995. *The minimalist program*. Cambridge, Mass.: MIT Press.

Chomsky, Noam. 2000. Minimalist inquiries. In *Step by step: Essays on minimalist syntax in honor of Howard Lasnik*, ed. by Roger Martin, David Michaels, and Juan Uriagereka, 89-155. Cambridge, Mass.: MIT Press.

Chomsky, Noam. 2001a. Derivation by phase. In *Ken Hale: A life in language*, ed. by Michael Kenstowicz, 1-52. Cambridge, Mass.: MIT Press.

Chomsky, Noam. 2001b. Beyond explanatory adequacy. In *MIT Working Papers in Linguistics* 20, 1-28. MITWPL, Department of Linguistics and Philosophy, MIT, Cambridge, Mass.

Chomsky, Noam, and Howard Lasnik. 1993. The theory of principles and parameters. In *Syntax: An international handbook of contemporary research*, Vol. 1, ed. Joachim Jacobs, Arnim von Stechow, Wolfgang Sternefeld, and Theo Vennemann, 506-569. Berlin: Walter de Gruyter.

Collins, Chris. 1997. *Local economy*. Cambridge, Mass.: MIT Press.

Dikken, Marcel den. in press. On the nature and distribution of successive cyclicity. In *Proceedings of NELS* 40, MIT.

Epstein, Samuel David. 1999. Un-principled syntax: The derivation of syntactic relations. In *Working Minimalism*, ed. by Samuel Epstein and Norbert Hornstein, 317-345. Cambridge, Mass: MIT Press.

Epstein, Samuel David, Erich Groat, Ruriko Kawashima, and Hisatsugu Kitahara. 1998. *A derivational approach to syntactic relations.* New York: Oxford University Press.

Epstein, Samuel David and T. Daniel Seely. 2006. *Derivations in Minimalism*. Cambridge: Cambridge University Press.

Fox, Danny, and David Pesetsky. 2005. Cyclic linearization of syntactic structure. *Theoretical Linguistics* 31: 1-45.

Franks, Steven. 1998. Clitics in Slavic. Presented at the Comparative Slavic Morphosyntax Workshop, Spencer Creek, Indiana, June 1998.

Franks, Steven, and Željko Bošković. 2001. An argument for multiple spell-out. *Linguistic Inquiry* 32: 174-183.

Gonzalez, Nora. 1988. *Object Raising in Spanish*. New York, NY.: Garland.

Groat, Erich, and John O'Neil. 1996. Spell-out at the LF interface. In *Minimal ideas*. ed. by Werner Abraham, Samuel David Epstein, Höskuldur Thráinsson, and C. Jan-Wouter Zwart, 113-139. Amsterdam: John Benjamins.

Grohmann, Kleanthes. 2003. *Prolific domains: on the anti-locality of movement dependencies*. Amsterdam: John Benjamins.

Hornstein, Norbert. 1999. Movement and control. *Linguistic Inquiry* 30: 69-96.

Hornstein, Norbert. 2001. *Move! A minimalist theory of construal*. Malden, Mass.: Blackwell.

Huang, C.-T. James. 1982. Logical relations in Chinese and the theory of grammar. Doctoral dissertation, MIT, Cambridge, Mass.

Kayne, Richard. 1994. *The antisymmetry of syntax*. Cambridge, Mass.: MIT Press.

Kayne, Richard. 2003. Pronouns and their antecedents. In *Derivation and explanation in the Minimalist Program*, ed. by Samuel D. Epstein and T. Daniel Seely, 133-166. Oxford: Blackwell.

Koizumi, Masatoshi. 1994. Layered specifiers. In *Proceedings of the North East Linguistic Society* 24, ed. by Mercè Gonzàlez, 255-269. GLSA, University of Massachusetts, Amherst.

Landau, Idan. 2007. EPP extensions. *Linguistic Inquiry* 38: 485-523.

Lasnik, Howard. 1995. Case and expletives revisited: On Greed and other human failings. *Linguistic Inquiry* 26: 615-633.

Lasnik, Howard. 2006. Conceptions of the Cycle. In Lisa Cheng and Norbert Corver (eds.) *Wh-Movement Moving On*. MIT Press, pp. 197-216.

Lasnik, Howard, and Mamoru Saito 1991. On the subject of infinitives. In *Papers from the 27th regional Meeting of the Chicago Linguistic Society Part One: The General Session*, ed. by Lise M. Dobrin, Lynn Nichols, and Rosa M. Rodriguez, 324-343. Chicago Linguistic Society, University of Chicago, Chicago, Ill.

Lebeaux, David. 1988. Language acquisition and the form of the grammar. Doctoral dissertation, University of Massachusetts, Amherst.

Martin, Roger. 1996. A minimalist theory of PRO and control. Doctoral dissertation, University of Connecticut.

Müller, Gereon. 2010. On Deriving CED Effects from the PIC. *Linguistic Inquiry* 41: 35-82.

Muysken, Pieter. 1982. Parametrizing the notion Head. *Journal of Linguistic Research* 2: 57-76.

Nissenbaum, Jon. 2000. Investigations of covert phrase movement. Doctoral dissertation, MIT, Cambridge, Mass.

Nunes, Jairo, 2004. *Linearization of chains and sideward movement*. Cambridge, Mass.: MIT Press.

Nunes, Jairo, and Juan Uriagereka.2000. Cyclicity and extraction domains. *Syntax* 3: 20-43.

Pesetsky, David, and Esther Torrego. 2001. T-to-C movement: Causes and consequences. In *Ken Hale: A life in language*, ed. Michael Kenstowicz, 355-426. Cambridge, Mass.: MIT Press.

Pesetsky, David, and Torrego, Esther. 2007. The syntax of valuation and the interpretability of features. In *Phrasal and Clausal Architecture: Syntactic derivation and interpretation. In honor of Joseph E. Emonds*, ed. by Simin Karimi, Vida Samiian and Wendy Wilkins, 262-294. Amsterdam: John Benjamins.

Rackowski, Andrea and Norvin Richards. 2005. Phase edge and extraction: A Tagalog case study, *Linguistic Inquiry* 36: 565–99.

Richards, Norvin. 2001. *Movement in language: Interactions and architectures*. Oxford: Oxford University Press.

Postal, Paul. 1974. *On raising: One rule of English grammar and its theoretical implications*. Cambridge,

Mass.: MIT Press.

Rizzi, Luigi. 1990. *Relativized minimality*. Cambridge, Mass.: MIT Press.

Rudin, Catherine. 1988. On multiple questions and multiple *wh*-fronting. *Natural Language and Linguistic Theory* 6: 445-501.

Simpson, Andrew. 2002. IP-raising, Tone Sandhi and the Creation of S-final Particles: Evidence for Cyclic Spell-Out. *Journal of East Asian Linguistics* 11: 67-99.

Starke, Michal. 2001. Move dissolves into Merge: A theory of locality. Doctoral dissertation, University of Geneva.

Stepanov, Arthur. 2001. *Cyclic domains in syntactic theory*. Doctoral dissertation, University of Connecticut.

Takahashi, Daiko. 1994. Minimality of movement. Doctoral dissertation, University of Connecticut.

Uriagereka, Juan. 1999. Multiple Spell-Out. In *Working minimalism*, ed. by Sam Epstein and Norbert Hornstein, 251-282. Cambridge, Mass.: MIT Press.