# Constraints on Grammatical Dependencies

Gereon Müller, Universität Leipzig March 21, 2020

## 1. Overview

Noam Chomsky is responsible for creating the field of research on grammatical theory in its present form, and he has also been material in providing frameworks for syntactic analysis, from early transformational grammar (see Chomsky (1957; 1975)) to current implementations of the minimalist program (see Chomsky (2001; 2008; 2013), Chomsky et al. (2019)). Against the background of these (and other) syntactic frameworks, a huge number of constraints on grammatical dependencies have been proposed over the years, most of them by other scholars, but quite a few of them also by Chomsky himself. Some of these constraints have become famous both within and outside of the field of linguistics, like the Subjacency Condition (Chomsky (1977)), the Empty Category Principle (ECP; Chomsky (1981)), and Principles A, B and C of the theory of binding (Chomsky (1981)). Other constraints initially suggested by Chomsky have a more troubled history, in the sense that they have variously been claimed to be empirically falsified, considered conceptually flawed, or simply consigned to oblivion. It is a main goal of the present contribution to argue that some of the constraints arguably belonging to this latter group can, upon closer scrutiny, in fact be shown to be at the heart of current syntactic theory, and qualify as "good" (in a sense to be made precise), whereas this does not necessarily hold for constraints in the former group. More specifically, based exclusively on a consideration of constructions involving displacement (movement) of syntactic categories, I will show that the A-over-A Principle (Chomsky (1964a)), the Superiority Condition (Chomsky (1973)), the Phase Impenetrability Condition (PIC; Chomsky (2001)) and the Strict Cycle Condition (SCC; Chomsky (1973; 1995)) together form the core of a comprehensive, beautiful theory of (relativized and rigid) locality, whereas constraints like the ECP have no place in current syntactic models.

## 2. Constraints

Sentences in natural languages have a hierarchical structure. This structure can be generated (or licensed) in various ways, by a structure-building component of grammar. One specific proposal that has been widely adopted in recent years is that syntactic structures are generated incrementally, from bottom to top, by a recursive operation of Merge (Chomsky (2001)), Hauser et al. (2002)). Merge can create basic structures like (1-a) by first combining the verb (V) with the object noun phrase (DP), which results in a VP, next combining an abstract functional head v (encoding agentivity) with the VP, which results in a v′ projection, and then combining v′ with the subject (DP), which results in a vP projection. In (1-a), all Merge operations are *external* since none of the two items that are put together is part of the other one before the operation applies. (1-bc) show that structure-building can also involve *internal* Merge, where one of the two items was part of the other one before the operation takes place. In (1-b), a functional head T(ense) is (externally) merged with the complex vP in (1-a), yielding T′; and internal Merge applies afterwards to T′ and the subject DP, leaving a trace (or a copy) in the original position (the trace t is coindexed here with the displaced DP for reasons of perspicuity). Internal Merge accounts for a universal property of natural languages, viz., that they all involve *displacement* (or movement): Items may occur in places which are clearly not their base positions. This already suggests itself in (1-b) (where $DP_1$ is introduced locally by v, not by the T element *will*); and it is even clearer in (1-c), where a TP that minimally differs from the one in (1-b) (by merging [$_{V}$ *like* ] with [$_{DP}$ *who(m)* ] instead of [$_{DP}$ *him* ]) is first externally merged with a C(omplementizer) head, and the resulting

C′ is then combined with the DP object, via internal Merge, thus bringing about so-called wh-movement and yielding a clause (CP). (1-c) can then be merged with a CP-embedding verb, and given further Merge operations in the matrix clause, this may lead to a sentence like (1-d). Alternatively, however, internal Merge will also apply to the embedded $DP_2$ and a matrix C′, as in (1-e), which yields non-clause bound displacement.

(1)
a. [$_{vP}$ [$_{DP}$ she ] [$_{V'}$ v [$_{VP}$ [$_{V}$ kiss ] [$_{DP}$ him ]]]]
b. [$_{TP}$ [$_{DP_1}$ she ] [$_{T'}$ [$_{T}$ will ] [$_{vP}$ $t_1$ [$_{V'}$ v [$_{VP}$ [$_{V}$ kiss ] [$_{DP}$ him ]]]]]]
c. [$_{CP}$ [$_{DP_2}$ who(m) ] [$_{C'}$ C [$_{TP}$ [$_{DP_1}$ she ] [$_{T'}$ [$_{T}$ will ] [$_{vP}$ $t_1$ [$_{V'}$ v [$_{VP}$ [$_{V}$ kiss ] $t_2$ ]]]]]]]
d. I wonder [$_{CP}$ [$_{DP_2}$ who(m) ] [$_{C'}$ C [$_{TP}$ [$_{DP_1}$ she ] [$_{T'}$ [$_{T}$ will ] [$_{vP}$ $t_1$ [$_{V'}$ v [$_{VP}$ [$_{V}$ kiss ] $t_2$ ]]]]]]]
e. [$_{CP}$ [$_{DP_2}$ Who(m) ] do you think [$_{CP}$ $t_2$ [$_{C'}$ C [$_{TP}$ [$_{DP_1}$ she ] [$_{T'}$ [$_{T}$ will ] [$_{vP}$ $t_1$ [$_{V'}$ v [$_{VP}$ [$_{V}$ kiss ] $t_2$ ]]]]]]]] ?

Clearly, there is no upper limit on the distance that an item like $DP_2$ in (1-e) can travel via internal Merge in a clause: Syntactic displacement is in principle unbounded; see, e.g., (2-a), with an additional CP separating *whom* from the position where it was externally merged (signalled by $t_1$). However, this is not yet the whole story. As was noted by Chomsky (1964a;b) early on, not everything is possible with displacement: There are restrictions that show that the structure-building component of syntax must be enriched by another component that contains syntactic constraints, so as to avoid overgeneration.[1] For instance, whereas (2-a) involves movement of an object across two complement clauses embedded by verbs, such movement becomes impossible if an intervening complement clause is embedded by a noun, as in (2-b) (where * signals ungrammaticality). Similarly, non-clause bound movement of subjects is much more restricted than non-clause bound movement of objects; see, e.g., (2-c) vs. (2-d) (= a version of (1-e) where C is not empty but realized by *that*).

(2)
a. Who(m)$_1$ do you think that Mary believes that she will kiss $t_1$ ?
b. *Which book$_1$ did John hear [$_{DP}$ a [$_{NP}$ [$_{N}$ rumour ] [$_{CP}$ that you had read $t_1$ ]]] ?
c. *Who$_1$ do you think [$_{CP}$ that $t_1$ left ] ?
d. Who(m)$_1$ do you think [$_{CP}$ that she will kiss $t_1$ ] ?

What data such as these show for movement can be generalized to all other grammatical dependencies relating two items in a sentence (like agreement, binding, case assignment, control, etc.): There must be constraints that restrict the power of pure structure-building in syntax. Given this state of affairs, the central question is what syntactic constraints look like.

Research over the last few decades has established a number of requirements that a syntactic constraint should meet in order to qualify as a plausible candidate for inclusion in the faculty of language. For the following discussion, I will adopt the criteria in (3).

(3)
a. Constraints are as simple and general as possible.
b. Constraints are not complex.
c. Constraints do not require massive search space.
d. Constraints are principles of efficient computation or interface conditions.

First, (3-a) is really just a manifestation of an uncontroversial requirement for all scien-

[1] Note that this conclusion is independent of the specific approach to syntactic structure-building chosen. Thus, if basic structure-building and displacement are not brought about by external and internal Merge, but rather by context-free phrase structure rules and designated transformations (which map syntactic trees onto other syntactic trees), respectively, as in classical transformational grammar (which of course formed the background to Chomsky's original considerations), the need for constraints arises in exactly the same way.

tific models: If there are two theories that account for the same set of data, the one is to be preferred that is superior with respect to simplicity, generality, and elegance. From this requirement, it follows that the number of Boolean operations in constraint definitions should be minimal (cf. Grimshaw (1998)). Furthermore, syntactic constraints should not mention specific information like, e.g., particular category (i.e., part of speech) features (or labels), or features characterizing arbitrary subclasses of the linguistic expressions that the constraint holds for. Evidently, construction-specific assumptions (let alone mentioning of individual lexical items) should be avoided in constraints. From this perspective, a constraint like the **that*-trace filter suggested in Chomsky & Lasnik (1977) to account for so-called complementizer-trace effects arising with movement of a subject wh-phrase from an embedded clause across a complementizer *that* to the matrix clause, as in (2-c), cannot be a good constraint.

Second, (3-b) presupposes that constraints on syntactic dependencies can be grouped into various classes determined by their complexity (see Müller & Sternefeld (2001), Graf (2013)). *Local derivational constraints* apply to individual syntactic operations (like Merge). *Local representational constraints* apply to syntactic output representations. *Global constraints* apply to whole derivations of sentences (i.e., successions of Merge and other operations); they correlate non-adjacent steps in the derivation (cf. Lakoff (1971); Chomsky's (1981) Projection Principle belongs in this class). *Transderivational constraints* are even more powerful since they require the construction of a reference set of competing syntactic derivations based on an initial input from which one (or more) derivation is picked out as optimal; this group includes the *Fewest Steps* condition selecting the derivation with the fewest number of operations that lead to convergence (Chomsky (2001)), the *Shortest Paths* condition selecting the derivation with the shortest movement paths (Chomsky (1995)), and the *Merge over Move* condition preferring a step of basic structure-building (external Merge) to a movement step (internal Merge), in a given local domain (the phase); cf. Chomsky (2000), Frampton & Gutmann (1999), Hornstein (2009). Finally, there is a fifth type that can be referred to as *translocal constraints*. In contrast to transderivational constraints, a translocal constraint compares competing output representations, and picks one as optimal; a canonical constraint of this type is the *Avoid Pronoun* principle that favours representations with non-overt pronouns over representations with overt pronouns (Chomsky (1981), Haegeman (1995)). Given this taxonomy, a natural assumption is that only *local* constraints can be good constraints.

Third, (3-c) states that in order to determine whether a (local) constraint is violated or not, it should ideally not be necessary to check structures of potentially unbounded size; rather, the smaller the accessible window of a complex syntactic object is for evaluation, the better the constraint satisfies this requirement. This requirement is violated, e.g., by Fox & Pesetsky's (2005) *Cyclic Linearization* constraint that demands a preservation of linear precedence relations established in a low syntactic domain (a VP, or a clause) in all higher domains of a potentially complex sentence.

Fourth, the requirements in (3-d) have been at the core of minimalist research (see, e.g., Chomsky (2005)). Principles of efficient computation in the sense of (3-d) are basically local derivational economy constraints which can be motivated on grammar-external grounds. In contrast, interface conditions in the sense of (3-d) are representational constraints on outputs which are not motivated syntax-internally, but result from demands imposed on syntactic representations by the phonological and semantic interfaces – Phonological Form (PF) and Logical Form (LF), respectively. Thus, according to this far-reaching requirement, constraints that rely on concepts without independent motivation in terms of either economy or interface conditions are excluded.

Based on the requirements in (3), the next section will discuss Ross's island constraints and the ECP, concluding that none of these constraints qualifies as good (in the technical sense just laid out).

## 3. Constraints That Do Not Meet The Requirements

### *3.1. Island Conditions*

Ross (1967) set out to develop a comprehensive system of so-called island constraints that severely restrict displacement operations in natural languages. However, it turns out that given the criteria in (3), none of the island constraints proposed in Ross (1967) (among them the *Complex NP Constraint*, the *Sentential Subject Constraint*, the *Coordinate Structure Constraint*, and the *Left Branch Condition*) qualifies as a good constraint. As a representative example, consider the Complex NP Constraint (CNPC) in (4), which blocks sentences like (5-a) (= (2-b)) in English (and many other languages), where wh-movement illegitimately takes place across both a CP ('sentence') and a DP ('noun phrase') with a lexical head (N), even though wh-movement from a complement CP is possible in principle (see (1-e), (2-a), (2-d)), as is wh-movement from an object DP (see (5-b)).

(4) *Complex NP Constraint* (CNPC):
No element contained in a sentence dominated by a noun phrase with a lexical head noun may be moved out of that noun phrase by a transformation.

(5) a. *Which book$_1$ did John hear [$_{DP}$ a [$_{NP}$ [$_N$ rumour ] [$_{CP}$ that you had read t$_1$ ]]] ?
b. Who$_1$ did you read [$_{DP}$ a [$_{NP}$ [$_N$ book ] [$_{PP}$ about t$_1$ ]]] ?

The CNPC is at variance with (3). For one thing, it is not maximally simple and general since it explicitly mentions category information (sentence:CP, noun phrase:DP, and lexical head noun:N) and thereby encodes a very specific construction. For another, the CNPC requires massive search space spanning the whole structure in (5); in fact, it is a two-node constraint that requires keeping track of a node crossed earlier. And finally, it is hard to see how (4) could be viewed as a principle of efficient computation, or as an interface constraint. Essentially the same conclusions can be drawn for the other constraints on movement dependencies in Ross (1967); but also for similar types of constraints like Chomsky's (1964a) *Wh-Island Condition* (according to which movement must not take place from indirect questions) and Chomsky's (1973) *Tensed-S Condition* and *Specified Subject Condition*, all of which describe specific constructions and thus wear their violations of (3-a) (and also of (3-d)) on their sleeves. And it goes without saying that the same kinds of consequences arise for proposals like Zaenen & Pinkham's (1976) *Post-Sentential Subject Extraction Constraint* or Kuno's (1973) *Clause Non-final Incomplete Constituent Constraint*. As noted by Chomsky (1982, 74-76), Ross's island constraints and several other constraints of this general type have been immensely important for the development of the field, but only insofar as they provide a "taxonomy of properties" that makes it possible to "rethink conditions, to see whether one could deal with the [...] phenomena [...] in a unified way" (Chomsky (2002, 129-130)) and to "explain [...] such constraints in terms of deeper and more natural principles from which their effects could be deduced" (Chomsky (1986b, 71-72)), and not as candidates for belonging to the faculty of language in their own right.[2]

[2] The *Subjacency Condition* (see Chomsky (1977; 1986a)) was introduced in order to derive effects covered by constraints like the CNPC and the Wh-Island Condition, by prohibiting movement that crosses two *bounding nodes*; the set of bounding nodes may vary from one language to another (see Rizzi (1982)). This two-node condition, while certainly an improvement over Ross' original constraints, can also easily be shown to violate

### *3.2. The Empty Category Principle*

The Empty Category Principle (ECP) was originally proposed in Chomsky (1981) (with refinements in Aoun, Hornstein & Sportiche (1981), Chomsky (1986a), Lasnik & Saito (1984; 1992), and other work). It is, without any doubt, one of the most successful and widely discussed constraints on syntactic dependencies ever.[3] Unlike, say, the CNPC, it is not a derivational local constraint (blocking a syntactic operation), but a representational local constraint (that filters out certain resulting structures). More specifically, it is a constraint on *traces*, which was made possible by the introduction of the trace theory of movement in the previous decade (cf. Wasow (1972), Chomsky (1973, 266-269), Fiengo (1977)). As shown in (6), the ECP in its canonical formulation is extremely complex.

(6) *Empty Category Principle* (ECP):
Every trace t is marked [+$\gamma$].
A trace t is marked [+$\gamma$] iff it is properly governed.
A trace t is properly governed iff (a) or (b) holds.
  a. t is antecedent-governed.
  A trace $\beta$ is antecedent-governed iff there is no barrier separating $\beta$ from its immediate chain antecedent $\alpha$ (and if there is no *wh*-phrase or overt complementizer in the C domain that intervenes between $\alpha$ and $\beta$).
  b. t is lexically governed.
  A trace $\beta$ is lexically governed iff it is in the m-command domain of a lexical category $\alpha$, and no barrier separates $\beta$ from $\alpha$.

The ECP applies to syntactic representations at the level of Logical Form (LF), after abstract operations like wh-movement of in-situ wh-phrases and deletion of *that* complementizers and of intermediate traces of arguments (subjects or objects) have applied. It demands that traces left by movement have to be marked [+$\gamma$]; a trace can be marked [+$\gamma$] only if it is properly governed; and it can be properly governed either via antecedent-government, or via lexical government. Given certain assumptions (in particular, the additional stipulation in brackets in (6-a), which is necessitated by the move from a single COMP position in Chomsky (1981) to a specifier head configuration SpecC-C in Chomsky (1986a)), the ECP can account for complementizer-trace effects with subject movement across a lexical complementizer as in (7-a) (= (2-c)), as opposed to displacement from a CP that does not have an overt complementizer; see (7-b). This kind of effect does not show up with object movement because object traces, unlike subject traces, are lexically governed; so they are always properly governed, hence marked [+$\gamma$], and can never violate the ECP.

(7) a. *Who$_1$ do you think [$_{CP}$ t$'_1$ [$_{C'}$ [$_C$ that ] [$_{TP}$ t$_1$ left ]]] ?
  b. Who$_1$ do you think [$_{CP}$ t$'_1$ [$_{C'}$ [$_C$ –] [$_{TP}$ t$_1$ left ]]] ?

The ECP in (6) also correctly predicts that the complementizer-trace effect disappears if the complementizer intervenes at a later stage of the derivation, and the original subject trace can be antecedent-governed by an intermediate trace, as in (8-a); this latter intermediate trace cannot be marked [+$\gamma$], but, not being a final member of a wh-movement chain, it can be deleted before the ECP is checked at LF. Similarly, assuming that adjunct (i.e., adverbial) traces can only be [$\gamma$]-marked at LF (whereas argument traces are [$\gamma$]-marked as

---

the requirements in (3).

[3] An extended google scholar search in early 2020 yields almost 7000 hits for articles or books containing "ECP" and "Chomsky".

soon as they arise in the derivation), the non-occurrence of a complementizer-trace effect in (8-b) can be accounted for by postulating complementizer deletion before LF. (Still, in other environments where adjunct movement has to cross a barrier, such movement is correctly excluded.)

(8) a. Who$_1$ do you believe [$_{CP}$ that Mary thinks [$_{CP}$ t$'_1$ [$_{C'}$ [$_C$ ] [$_{TP}$ t$_1$ left ]]]] ?
b. Why$_1$ do you think [$_{CP}$ t$'_1$ [$_{C'}$ [$_C$ that ] [$_{TP}$ Mary left t$_1$ ]]] ?

The ECP also covers many so-called superiority effects, as in (9-b) vs. (9-a), assuming that there is LF wh-movement of the in-situ wh-phrase in multiple questions in English (which gives rise to a subject LF trace $t_1$ in (9-b) that is not properly governed, in contrast to the object LF trace $t_2$ in (9-a)).

(9) a. I wonder [$_{CP}$ who$_1$ t$_1$ bought what$_2$ ] $\rightarrow$ LF: [$_{CP}$ wh$_2$ wh$_1$ C t$_1$ ... $\boxed{t_2}$ ]
b. *I wonder [$_{CP}$ what$_2$ who$_1$ bought t$_2$ ] $\rightarrow$ LF: [$_{CP}$ wh$_1$ wh$_2$ C $\boxed{t_1}$ ... t$_2$ ]

In addition, subsequent research has identified many more ECP effects of diverse provenance, and in many different languages, exhibiting asymmetries between subject and object movement on the one hand, and between argument and adjunct movement on the other. However, despite its considerable empirical coverage, the ECP is not a good constraint. First, as should be clear from (6), the ECP has a highly complex inner structure, involving a huge number of auxiliary concepts and a disjunction at its core (antecedent-government or lexical government); so it certainly does not look simple and general. This conclusion is reinforced by the fact that the constraint talks about a specific type of empty category (traces left by movement), rather than empty categories in general (including the *pro* postulated as an empty subject for seemingly subject-less sentences like *Ha parlato* ('has spoken') in Italian, and the PRO postulated as an empty subject of infinitives in control structures like *She tries to win*). And second, the ECP cannot naturally be conceived of either as a principle contributing to efficient computation, or a condition imposed by the LF interface (the fact that it is *formulated* as an LF constraint notwithstanding). A systematic account of ECP effects in terms of constraints that meet the requirements in (3) is still outstanding.[4]

## 4. Constraints That Meet The Requirements

In contrast to the constraints discussed in the previous section, there are two pairs of constraints introduced by Noam Chomsky which are at the very heart of current approaches to locality in syntactic dependencies: on the one hand, the A-over-A Principle and the Superiority Condition (relativized locality), and on the other hand, the Phase Impenetrability Condition and the Strict Cycle Condition (rigid locality). I address the two concepts in turn.

### *4.1. Relativized Locality Constraints*

#### *4.1.1. The A-over-A Principle*

The A-over-A Principle is arguably the first substantial locality constraint on grammatical dependencies that has ever been proposed; see Chomsky (1964a;b); the formulation in (10) essentially follows Chomsky (1973, 235) (also cf. Ross (1967, 10), Gazdar (1982, 176-177), Riemsdijk & Williams (1986, 20) for slightly different formulations).

[4] A general *Anti-Locality Constraint* (see Bošković (1994; 2014), Grohmann (2000; 2003), Grohmann & Nevins (2004), Abels (2003; 2012), Erlewine (2016), Pesetsky (2016)) can block subject movement as too short where object movement in the same context is sufficiently non-local, and might thus ultimately be a candidate for this role.

(10) *A-over-A Principle*:
If a transformation applies to a structure of the form $[_{\alpha}$ ... $[_{A}$ ... $[_{A}$ ... ] ... ] ...], where $\alpha$ is a cyclic node, then it must be so interpreted as to apply to the higher phrase of the type A.

The standard interpretation (and also the original one adopted in Chomsky (1964b)) is that "A" in (10) is a variable over category features, i.e., labels. The constraint then states that movement of some category $\gamma$ to some landing site $\Delta$ (in fact, any grammatical dependency involving $\gamma$ and $\Delta$) is impossible if there another category $\beta$ with the same category feature that dominates $\gamma$ (in this case, $\beta$ is *closer* to $\Delta$ than $\gamma$ because the path from $\beta$ to $\Delta$ is necessarily shorter than the path from $\gamma$ to $\Delta$). The A-over-A Principle is a local derivational constraint. It predicts that, within a given cyclic domain, a DP cannot be extracted from another DP; a VP cannot be extracted from a higher VP, and so on.[5] As such, the constraint makes a number of non-trivial correct predictions; however, as pointed out by Ross (1967), in the (label-based) interpretation standardly adopted, the A-over-A Principle leads to both undergeneration and overgeneration. The second problem may make the A-over-A Principle look less plausible but is not fatal (after all, the data that the A-over-A principle cannot exclude even though one might intuitively think that it should – like, e.g., cases of illicit AP extraction from DP, as in **How intelligent do you have a sister?* – can always be excluded by some other constraint(s) in the grammar). The first problem is potentially severe, though. Examples like (5-b) show that a DP can in principle be moved out of another DP. Examples like those in (11-ab) (from English and German, respectively) indicate that a VP can undergo topicalization across another VP, and that a CP can be topicalized over another CP that initially dominates it.

(11) a. $[_{VP_2}$ Give a book to Mary ] I don't $[_{VP_1}$ think $[_{CP}$ that she will $t_2$ ]]
b. $[_{CP_2}$ Dass er nett ist ] hat Fritz behauptet $[_{CP_1}$ dass Maria gesagt hat $t_2$ ]
that he nice is has Fritz$_{nom}$ claimed that Maria$_{nom}$ said has

The wellformedness of these kinds of examples strongly suggests that the A-over-A Principle cannot be maintained in the interpretation given to "A" so far, viz., as a category label. However, following Ross (1967), the vast majority of the field seems to have reached the much more far-reaching conclusion that the A-over-A Principle should be abandoned in toto. In line with this, most textbooks on syntactic theory do not even mention this constraint anymore. Where it shows up, it is typically characterized as having been falsified by Ross (see, e.g., Riemsdijk & Williams (1986)). Very few textbooks introduce the constraint from a more neutral point of view (see Freidin (1992) for one exception), and even fewer have something positive to say about it (but cf. Hornstein et al. (2005), who mention the A-over-A Principle in passing).[6]

However, the obvious alternative to abandoning the constraint is of course to revise the assumption that "A" in (10) stands for a category label. As a matter of fact, this is exactly what Chomsky (1973, 235) does: He points out that it is possible to assume that the A-over-A Principle in (10) "does not establish an absolute prohibition against transformations that extract a phrase of type A from a more inclusive phrase of type A. Rather, it states that if a

[5] From a contemporary point of view based on incremental structure-building via Merge, the most restrictive concept of a *cyclic domain* is the current root in a given derivation.
[6] In view of Ross's counter-arguments, Postal (2012) takes this constraint to be a "failed principle", a "bankrupt idea", and criticizes Chomsky for not abandoning the A-over-A Principle after Ross (1967). As we will see momentarily, both Postal's assessment of the status of the A-over-A Principle and his criticism of Chomsky based on it are misguided.

transformational rule is nonspecific with respect to the configuration defined, it will be interpreted in such a way as to satisfy the condition." As Chomsky also notes, "alternatively, one might interpret the A-over-A constraint as legislating against any rule that extracts a phrase of type A from a more inclusive phrase A," which is the standard understanding facing empirical counter-evidence. Importantly, Chomsky's (1973) actual conclusion already is that "the former interpretation [...] is perhaps more natural", and that he "will adopt it tentatively here, for this and other conditions to be discussed." What is more, this version of the A-over-A Principle is also adopted by Bresnan (1976): The item that can be subjected to a given transformation should be as inclusive as possible, independently of category labels (whereas an even more inclusive item that cannot be subjected to the same transformation because it does not fit its structural description can be ignored, even if it shares the same category label). Thus, under the revised understanding of "A" in (10) as the feature that is involved in the operation, the A-over-A Principle predicts that a wh-phrase cannot be moved out of a wh-phrase (but can move out of a non-wh phrase), a topic-marked phrase cannot move out of another topic-marked phrase, and so on. Let us now adopt a more contemporary approach according to which all syntactic operations are driven by designated features (cf. Chomsky (1995; 2001), Pesetsky & Torrego (2006), Urk (2015), Collins & Stabler (2016) and Georgi (2017), among many others):[7] There are structure-building features (rendered here as [•F•]) that trigger movement and basic structure-building (internal and external Merge), and there are probe features (rendered as [∗F∗]) that trigger Agree operations as they underlie grammatical dependencies like case assignment, agreement, binding, and control (cf. Assmann et al. (2015) and references cited there for the notation). As far as movement is concerned, the A-over-A Principle can then be given the reformulation in (12), which is completely faithful to the reinterpretation of (10) along the lines of Chomsky (1973) and Bresnan (1976).

(12) *A-over-A Principle* (updated version):
In a structure $\alpha_{[\bullet F\bullet]}$ ... [ ... $\beta_{[F]}$ ... $\gamma_{[F]}$ ... ] ..., $\gamma$ cannot move to [•F•] if $\beta$ dominates $\gamma$.

Incidentally, this version of the A-over-A Principle corresponds to a relativized locality constraint that has widely been proposed in the more recent literature (cf. Koizumi (1995), Fukui (1997), Kitahara (1997), Müller (1998), Sauerland (1999), Fitzpatrick (2002), Vicente (2007), Heck (2008), and Pesetsky (2013)), typically without reference to either Chomsky (1973) or Bresnan (1976). (12) provides a straightforward explanation for an otherwise mysterious restriction on the movement of remnant categories (i.e., categories from which extraction has taken place): Remnant movement of a certain type is impossible if the movement operation that has created the remnant is of the same type (cf. Takano (1994), Müller (2014)). This presupposes that different kinds of movement can be distinguished. For present purposes, we can confine attention to *scrambling* (see Ross (1967)), which is a movement type creating word order variation in free word order languages by moving a DP to a specifier of v (Specv); *wh-movement*, which moves a wh-phrase to a specifier of C (SpecC; see (1-c)); and *topicalization*, which is movement of a non-wh-phrase to SpecC. Given this assumption, the effect that the A-over-A Principle in (12) has on remnant movement is shown in (13-a) for DP scrambling from VP co-occurring with remnant VP scrambling in German (where VP topicalization would be perfect in the same context), in (13-b) for DP scrambling from CP co-occurring with remnant CP scrambling in Japanese, and in (13-c) for wh-movement from a wh-phrase co-occurring with remnant wh-movement of that latter, more inclusive wh-phrase in English.

[7] However, see also Chomsky et al. (2019), Safir (2019) and literature cited there for the opposite view, essentially the "Affect-$\alpha$" view of standard government-and-binding theory (Lasnik & Saito (1992)).

(13) a. *dass [$_{VP_2}$ t$_1$ zu lesen ] [$_{DP_1}$ das Buch ] keiner t$_2$ versucht hat
that to read the book$_{acc}$ no-one tried has

b. *[$_{CP_2}$ Mary-ga t$_1$ yonda-to ] [$_{DP_1}$ sono hon-o ] John-ga t$_2$ itta (koto)
Mary$_{nom}$ read-COMP that book$_{acc}$ John$_{nom}$ said fact

c. *[$_{DP_2}$ Which picture of t$_1$ ] do you wonder [$_{CP}$ who$_1$ she likes t$_2$ ] ?

Given that scrambling and wh-movement involve different designated features on the attracting head and the moved item (like [•Σ•]/[Σ] for scrambling and [•wh•]/[wh] for wh-movement), the A-over-A Principle requires XP$_2$ to move first in (13-abc), so that subsequent movement of XP$_1$ from the XP$_2$ in the derived position to a lower position will have to violate a number of other constraints (in particular, the Condition on Extraction Domain and the Strict Cycle Condition; see below).

Finally, note that this version of the A-over-A Principle meets all the requirements in (3). It is simple, general, and not complex. It is compatible with a larger search space, but it does not need it to be properly evaluated in any given derivation. And since it is essentially a minimality requirement that brings about a selection of the closest matching item, it can plausibly be viewed as a constraint contributing to efficient computation.

#### *4.1.2. The Superiority Condition*

The original formulation of the Superiority Condition is given in Chomsky (1973, 246).

(14) *Superiority Condition*:
No rule can involve, X, Y in the structure
... X ... [$_\alpha$ ... Z ... -WYZ ... ] ...
where the rule applies ambiguously to Z and Y and Z is superior to Y.

In current terms, a category is superior to another one if it asymmetrically c-commands it, i.e., if it shows up higher in the structure.[8] Furthermore, the condition that a rule applies ambiguously to two items captures exactly the same scenario that we have seen with the revised A-over-A Principle: Both items (Z and Y in (14)) can in principle be affected by the (movement) operation. (14) then demands that the higher, c-commanding one is to be affected by the transformation (or to be affected *first* if the transformation can apply more than once). Like the A-over-A Principle, the Superiority Condition can be given a fully faithful update, as in (15).

(15) *Superiority Condition* (updated version):
In a structure $\alpha_{[\bullet F \bullet]}$ ... [ ... $\beta_{[F]}$ ... $\gamma_{[F]}$ ... ] ..., $\gamma$ cannot move to [•F•] if $\beta$ c-commands $\gamma$.

The Superiority Condition straightforwardly accounts for wh-movement asymmetries in multiple questions, as in (9) in English, without recourse to either the ECP or abstract LF movement: A subject wh-phrase c-commands an object wh-phrase and thus blocks its movement. Furthermore, unlike the ECP, the Superiority Condition accounts for similar intervention effects with two wh-objects, as in (16) (see Pesetsky (1982, 602ff)).

(16) a. Whom$_1$ did John persuade t$_1$ [$_{CP}$ to visit whom$_2$ ] ?
b. *Whom$_2$ did John persuade whom$_1$ [$_{CP}$ to visit t$_2$ ] ?

The constraint can also cover wh-island effects as in (17), where wh-movement of *how* ille-

---

[8] A category $\alpha$ c-commands another category $\beta$ iff the next category containing $\alpha$ also contains $\beta$, and neither contains the other; see Reinhart (1976). $\alpha$ asymmetrically c-commands $\beta$ if $\alpha$ c-commands $\beta$, and $\beta$ does not c-command $\alpha$.

gitimately takes place from a wh-clause; thus, invoking a more specificic constraint like the Wh-Island Condition or the Subjacency Condition is not necessary to exclude these cases.

(17) *How$_1$ does [$_{TP_3}$ she know [$_{CP}$ [$_{DP_2}$ which car ] [$_{TP_4}$ Mary fixed t$_2$ t$_1$ ]]] ?

In contrast, the Superiority Condition has nothing to say about topicalization from wh-islands, given that different movement-related features are involved ([top] vs. [wh]). However, as noted by Fanselow (1987), the relevant examples are in fact well formed in German (unlike cases of wh-movement from wh-clauses); so this is a welcome consequence.

Like the A-over-A Principle, the Superiority Condition qualifies as a good constraint. First, it is simple and general. It is worth pointing out that, although it is usually treated as a constraint purely on wh-movement, there is nothing in the formulation of the Superiority Condition that would support the idea that it is confined to wh-environments.[9] Second, it is a local derivational constraint, and hence not complex. Third, it is compatible with, but does not presuppose, a large search space. And fourth, it implements a minimality requirement that lends itself to a classification as a principle of efficient computation.

Evidently, the A-over-A Principle in (12) and the Superiority Condition in (15) are two sides of the same coin; both constraints demand movement of the closest item, and the only difference is that closeness is defined in terms of domination in the first case, and in terms of c-command in the second. It thus makes sense to subsume the two constraints under a single, generalized *Minimal Link Condition* (MLC; see Chomsky (1995; 2001; 2008), and Takano (1994), Kitahara (1997), Müller (1998), Fitzpatrick (2002), Rackowski & Richards (2005), and Medeiros (2008), among many others).

### *4.2. Rigid Locality Constraints*

The two constraints discussed in the previous section are *relativized* locality constraints since movement options for some item $\gamma$ are not restricted to a fixed domain; rather, they are determined by the presence or absence of an intervening item $\beta$ that is closer to the landing site. In contrast, *rigid* locality constraints on syntactic dependencies identify intrinsic, invariant properties of domains that make them (in-) accessible for syntactic operations.

#### *4.2.1. The Strict Cycle Condition*

The Strict Cycle Condition in (18) has been proposed for syntactic dependencies in Chomsky (1973, 243).

(18) *Strict Cycle Condition* (SCC):
No rule can apply to a domain dominated by a cyclic node *A* in such a way as to affect solely a proper subdomain of *A* dominated by a node *B* which is also a cyclic node.

The Strict Cycle Condition is arguably the single most important constraint in every derivational theory of grammar; see Chomsky (1995; 2001; 2004; 2008; 2019), Jacobson & Neubauer (1974), Perlmutter & Soames (1979)) Pullum (1992; 1979), Collins (1997), Kitahara (1997), Bošković & Lasnik (1999), Freidin (1999), Müller (2011), Abels (2012), Collins & Stabler (2016), and Gallego (2020), among many others, for discussion and alternative formulations.[10] This constraint ensures that once the derivation has created a certain

[9] Indeed, explicitly general minimality constraints that are virtually indistinguishable from (15) have been proposed subsequently, by Fanselow (1990; 1991) and by Ferguson (1993) and Ferguson & Groat (1994). Note incidentally that Rizzi's (1990) Relativized Minimality, which at first sight may seem similar to (15), is very different both conceptually and in its empirical consequences; see Müller (2011, 61-64).

[10] Versions of the constraint include the *Extension Condition* in Chomsky (1995), and the *No Tampering Condition* in Chomsky (2008).

structure, it is impossible to exclusively modify an embedded part of this structure. The degree of strictness of the constraint depends on how "cyclic node" is defined. According to the most restrictive concept, every projection is a cyclic node (also cf. footnote 5). On this view, every legitimate operation must involve the current root. Among many other things, the Strict Cycle Condition is indispensable in a derivational account of freezing effects (Wexler & Culicover (1980), Browning (1991), Lohndal (2010)), as in (19-a) (with DP raising to subject in the passive interacting with wh-movement from DP in English) and (19-b) (with VP topicalization interacting with wh-movement from VP in German).

(19) a. *Who$_1$ was [$_{\text{DP}_2}$ a picture of t$_1$ ] painted t$_2$ by Mary ?
b. *Was$_1$ denkst du [$_{\text{CP}}$ [$_{\text{VP}_2}$ t$_1$ gelesen ] hat keiner t$_2$ ] ?
what$_{acc}$ think you read has no-one$_{acc}$

Given that extraction from XP is possible only if XP is a complement (cf. the *Condition on Extraction Domain*; see Huang (1982), Chomsky (1986a), Cinque (1990)), the illformedness of (19-ab) can be derived if movement of DP$_2$ and VP$_2$ precedes extraction of DP$_1$ (because DP$_2$/VP$_2$ occupies a specifier when DP$_1$ extraction takes place); but the reverse, counter-cyclic sequence of movement operations where DP$_1$ extraction takes place when DP$_2$/VP$_2$ is still in situ, in a complement position, must also be excluded, and this is accomplished by the Strict Cycle Condition. Similarly, if it is assumed that the embedded wh-phrase in SpecC is responsible for the wh-island effect in a sentence like (17), the Strict Cycle Condition is needed to ensure that it is already present when movement of the other wh-phrase to the matrix domain takes place.

More generally, as observed by McCawley (1984; 1998), the Strict Cycle Condition predicts *orders* among syntactic operations. Given a very narrow concept of cyclic domains (the minimal projection), the constraint can thus resolve what would otherwise be indeterminacies in rule application. As a case in point, consider the dual role of (transitive) v heads. Under standard assumptions, v introduces an agentive (external) argument (i.e., it has a structure-building feature [•D•] triggering Merge of a DP with v′), and it can assign accusative case as an instance of Agree (i.e., there is a valued probe feature [∗acc∗] in addition). It is a priori unclear which operation has to take place first (with radically different consequences potentially arising; see Müller (2004)); but the Strict Cycle Condition working on minimal cyclic domains resolves the order in favour of Agree preceding Merge: Agree operates under c-command (cf. footnote 8); thus, Agree of v and an object DP in the VP affects solely the v′ domain. Hence, if such v′-based Agree were to take place after structure-building via Merge of the subject DP, i.e., in the presence of a higher cyclic domain vP, the Strict Cycle Condition would be violated; therefore Agree must precede Merge.

Sometimes a weaker version of the Strict Cycle Condition is adopted. For instance, Safir (2019, 292) proposes a *Peak Novelty Condition* that permits operations which are not *massively* counter-cyclic (i.e., reasonably close to the current root domain). Now and then, the concept of *Featural Cyclicity* has been pursued (cf. Richards (2001) and Preminger (2018), among others); here the only requirement is that active features that can trigger operations must do so as soon as possible; i.e., they cannot wait and become embedded by further structure-building. From a historical perspective, Featural Cyclicity is similar to the *Cyclic Principle* of classic transformational grammar (cf., e.g., Perlmutter & Soames (1979)): When two operations can be carried out, where one applies to the cyclic domain D$_x$ and the other applies to the cyclic domain D$_{x-1}$ included in D$_x$, then the latter is applied first. Note that the Cyclic Principle and the Strict Cycle Condition are far from the same thing: A rule applying in a higher cycle may change the context for the application of a rule in a lower cycle, such that, e.g., the second rule can now apply (as part of the higher cycle). This interaction is excluded by the Strict Cycle Condition, but not by the Cyclic Principle.

The different predictions of the Strict Cycle Condition on the one hand and Featural Cyclicity and the Cyclic Principle on the other are striking. Numerous counter-cyclic operations that have been proposed are incompatible with the former constraint but compatible with the latter ones, among them late adjunction of adjuncts (Lebeaux (1988), Epstein et al. (1998) vs. Chomsky (2004)); wholesale later merger (Takahashi & Hulsey (2009), Urk (2015) vs. Chomsky (2019)); tucking in (Richards (2001)); head movement as adjunction (cf. Chomsky (1995), Matushansky (2006), Georgi & Müller (2010) for alternatives); and feature inheritance (Chomsky (2008), Richards (2007)).

Like the A-over-A Principle and the Superiority Condition, the Strict Cycle Condition qualifies as a good constraint according to (3). It is simple and not restricted to particular categories; it is not complex; it brings about a massive reduction of accessible space for items that can trigger operations; and by doing this and ensuring maximal stability of linguistic objects created in the course of the derivation, it looks like a good candidate for a principle of efficient computation.

#### *4.2.2. The Phase Impenetrability Condition*

Another fundamental constraint imposing rigid locality on operations is the Phase Impenetrability Condition (PIC) introduced in Chomsky (2000, 108) and Chomsky (2001, 13), and adopted in much subsequent minimalist work. The more restrictive version of the two PICs Chomsky considers is given in (20).[11]

(20) *Phase Impenetrability Condition* (PIC):
The domain of a head X of a phase XP is not accessible to operations outside XP; only X and its edge are accessible to such operations.

The edge of a head X is the residue outside of X′; it comprises specifiers of X (and adjuncts to XP). Chomsky (2000; 2001; 2008) proposes that phases are the clause (CP) and the predicate with its core arguments (vP). It follows from the PIC that all material that needs to be accessible outside the current phase must occur in the edge domain of the phase or on the phase head. For non-local movement operations ending beyond the current phase (as they can be found in most languages), this requires an intermediate step to a specifier of the phase head; for non-local agreement operations (as they can be found in some languages, like Hindi, Tsez, Itelmen, and Hinuq), the PIC may suggest a concept like cyclic Agree (Legate (2005)) or piggybacking of agreement on movement (Polinsky & Potsdam (2001)).[12]

Crucially, the PIC can act as a locality constraint on movement and have substantial effects here, but only if it is accompanied by a theory of what can, and what cannot, undergo movement to the phase edge. Pertinent suggestions include specific restrictions on the insertion of edge features as triggers for intermediate movement steps (see Müller (2011)) and, more recently, an Anti-Locality Constraint that prohibits items from reaching the phase edge that are too close to it (see the references in footnote 4).

Like the Strict Cycle Condition, the PIC qualifies as a good constraint in the sense of (3). It is a local constraint, hence, not complex. It clearly leads to a massive reduction of search space. It was explicitly introduced by Chomsky (2000; 2001) as a principle of efficient computation. And it is also simple and general. In particular, it can be observed that even

[11] Koster's (1978) *Bounding Condition* and Riemsdijk's (1978) *Head Constraint* can be viewed as predecessors.
[12] Alternatively, Chomsky et al. (2019) suggest that the complement of a phase head might remain accessible throughout; it cannot be modified, but goal features in this domain can be accessed for Agree relations.

though it singles out CPs and vPs as phases, this does not have to be stipulated but follows because these two categories form a natural class defined by the independently motivated feature “propositional”.[13]

As with the two relativized locality constraints, the two rigid locality constraints highlighted in this section form two sides of the same coin. The Strict Cycle Condition and the Phase Impenetrability Condition both reduce search space, but whereas the Strict Cycle Condition imposes restrictions on the position of an item that triggers an operation ([•F•], [∗F∗]), the PIC imposes restrictions on the position of an item that is sought by the trigger ([F]). The small size of the accessible window in syntactic derivations resulting from these two constraints, and the constant change of this window in the course of the derivation, are illustrated in (21) (where phases are underlined).

(21) *Search space reduction*:

a. $\underbrace{\overbrace{[_{\text{YP}} \ldots \text{Y}}^{\text{SCC}}\ [_{\underline{\text{XP}}} \ldots [_{\text{X}'} \text{X}}_{\text{PIC}}\ [_{\text{WP}} \ldots \text{W}\ [_{\underline{\text{UP}}} \ldots \text{U} \ldots ]]]]]$

b. $\underbrace{\overbrace{[_{\underline{\text{ZP}}} \ldots \text{Z}}^{\text{SCC}}\ [_{\text{YP}} \ldots \text{Y}\ [_{\underline{\text{XP}}} \ldots [_{\text{X}'} \text{X}}_{\text{PIC}}\ [_{\text{WP}} \ldots \text{W}\ [_{\underline{\text{UP}}} \ldots \text{U} \ldots ]]]]]$

To conclude, Chomsky’s A-over-A Principle, Superiority Condition, Strict Cycle Condition and Phase Impenetrability Condition are four constraints on grammatical dependencies that have widely been adopted in recent work on syntax, and that can be classified as good constraints according to natural criteria related to simplicity, generality, complexity, domain dependence, and efficient computation. In a nutshell, these constraints ensure that all grammatical dependencies obey minimality in syntactic structures, and that the active parts of syntactic structures that need to be considered are extremely small to begin with.

More generally, Chomsky’s legacy in this domain emerges as enormous: He was the first to see that there need to be constraints on grammatical dependencies; he has insisted that constraints should not be surface-oriented statements that merely lead to descriptive adequacy, but should instead be conceived of as abstract, simple principles that meet criteria like those in (3); and he has proposed a set of such constraints himself over the decades that form a core part of synctactic theory to this day.

## References


Abels, Klaus (2003): Successive Cyclicity, Anti-Locality, and Adposition Stranding. PhD thesis, University of Connecticut, Storrs, Connecticut.

Abels, Klaus (2012): *Phases. An Essay on Cyclicity in Syntax*. Vol. 543 of *Linguistische Arbeiten*, De Gruyter, Berlin.

Aoun, Joseph, Norbert Hornstein & Dominique Sportiche (1981): Aspects of Wide Scope Interpretation, *Journal of Linguistic Research* 1, 69–95.

Assmann, Anke, Doreen Georgi, Fabian Heck, Gereon Müller & Philipp Weisser (2015): Ergatives Move Too Early. On an Instance of Opacity in Syntax, *Syntax* 18, 343–387.


[13] Basically the same conclusion holds if *all* XPs are viewed as phases (Müller (2011)); arguably, some additional effort is required from this perspective to motivate proposals where only CP is a phase (Keine (2016)); where DP is a phase in addition to CP and vP (Matushansky (2005), Kramer (2007)); where TP can be a phase (Richards (2011)); or where phase status is more flexible (Bobaljik & Wurmbrand (2003), Marušič (2005), den Dikken (2007), Gallego (2007)).

Bobaljik, Jonathan & Susanne Wurmbrand (2003): Relativized Phases. Ms., University of Connecticut, Storrs.

Bošković, Željko (1994): D-Structure, Theta-Criterion, and Movement into Theta Positions, *Linguistic Analysis* 24, 247–286.

Bošković, Željko (2014): Now I'm a Phase, Now I'm Not a Phase: On the Variability of Phases with Extraction and Ellipsis, *Linguistic Inquiry* 45, 27–89.

Bošković, Željko & Howard Lasnik (1999): How Strict is the Cycle?, *Linguistic Inquiry* 20, 691–703.

Bresnan, Joan (1976): On the Form and Functioning of Transformations, *Linguistic Inquiry* 7, 3–40.

Browning, Marguerite (1991): Bounding Conditions on Representations, *Linguistic Inquiry* 22, 541–562.

Chomsky, Noam (1957): *Syntactic Structures*. Mouton, The Hague and Paris.

Chomsky, Noam (1964a): *Current Issues in Linguistic Theory*. Mouton, The Hague.

Chomsky, Noam (1964b): The Logical Basis of Linguistic Theory. In: H. Lunt, ed., *Proceedings of the Ninth International Congress of Linguists, Cambridge, Mass.*. Mouton and Co., The Hague.

Chomsky, Noam (1973): Conditions on Transformations. In: S. Anderson & P. Kiparsky, eds., *A Festschrift for Morris Halle*. Academic Press, New York, pp. 232–286.

Chomsky, Noam (1975): *The Logical Structure of Linguistic Theory*. Plenum Press, New York.

Chomsky, Noam (1977): On Wh-Movement. In: P. Culicover, T. Wasow & A. Akmajian, eds., *Formal Syntax*. Academic Press, New York, pp. 71–132.

Chomsky, Noam (1981): *Lectures on Government and Binding*. Foris, Dordrecht.

Chomsky, Noam (1982): *The Generative Enterprise. A Discussion with Riny Huybregts and Henk van Riemsdijk*. Foris Publications, Dordrecht.

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

Chomsky, Noam (1986b): *Knowledge of Language*. Praeger, New York.

Chomsky, Noam (1995): *The Minimalist Program*. MIT Press, Cambridge, Mass.

Chomsky, Noam (2000): Minimalist Inquiries: The Framework. In: R. Martin, D. Michaels & J. Uriagereka, eds., *Step by Step*. MIT Press, Cambridge, Mass., pp. 89–155.

Chomsky, Noam (2001): Derivation by Phase. In: M. Kenstowicz, ed., *Ken Hale. A Life in Language*. MIT Press, Cambridge, Mass., pp. 1–52.

Chomsky, Noam (2002): *On Nature and Language*. Cambridge University Press, Cambridge.

Chomsky, Noam (2004): Beyond Explanatory Adequacy. In: A. Belletti, ed., *Structures and Beyond. The Cartography of Syntactic Structures*. Vol. 3, Oxford University Press, Oxford, pp. 104–131.

Chomsky, Noam (2005): Three Factors in Language Design, *Linguistic Inquiry* 36, 1–22.

Chomsky, Noam (2008): On Phases. In: R. Freidin, C. Otero & M. L. Zubizarreta, eds., *Foundational Issues in Linguistic Theory*. MIT Press, Cambridge, Mass., pp. 133–166.

Chomsky, Noam (2013): Problems of Projection, *Lingua* 130, 33–49.

Chomsky, Noam (2019): Lectures. UCLA & MIT.

Chomsky, Noam & Howard Lasnik (1977): Filters and Control, *Linguistic Inquiry* 8, 425–504.

Chomsky, Noam, Ángel Gallego & Dennis Ott (2019): Generative Grammar and the Faculty of Language: Insights, Questions, and Challenges, *Catalan Journal of Linguistics* Special Issue, 229–261.

Cinque, Guglielmo (1990): *Types of A-bar Dependencies*. MIT Press, Cambridge, Mass.

Collins, Chris (1997): *Local Economy*. MIT Press, Cambridge, Mass.
Collins, Chris & Edward Stabler (2016): A Formalization of Minimalist Syntax, *Syntax* 19(1), 43–78.
den Dikken, Marcel (2007): Phase Extension: Contours of a Theory of the Role of Head Movement in Phrasal Extraction, *Theoretical Linguistics* 33, 1–41.
Epstein, Sam, Erich Groat, Ruriko Kawashima & Hisatsugu Kitahara (1998): *A Derivational Approach to Syntactic Relations*. Oxford University Press, Oxford and New York.
Erlewine, Michael (2016): Anti-Locality and Optimality in Kaqchikel Agent Focus, *Natural Language and Linguistic Theory* 34, 429–479.
Fanselow, Gisbert (1987): *Konfigurationalität*. Narr, Tübingen.
Fanselow, Gisbert (1990): On the Nature of Proper Government and Syntactic Barriers. In: W. Abraham, W. Kosmeijer & E. Reuland, eds., *Issues in Germanic Syntax*. Mouton de Gruyter, Berlin, pp. 33–48.
Fanselow, Gisbert (1991): Minimale Syntax. Habilitation thesis, Universität Passau.
Ferguson, K. Scott (1993): Notes on the Shortest Move Metric and Object Checking, *Harvard Working Papers in Linguistics* 3, 65–80.
Ferguson, Scott & Erich Groat (1994): Defining 'Shortest Move'. Ms., Harvard University.
Fiengo, Robert (1977): On Trace Theory, *Linguistic Inquiry* 8, 35–61.
Fitzpatrick, Justin (2002): On Minimalist Approaches to the Locality of Movement, *Linguistic Inquiry* 33, 443–463.
Fox, Danny & David Pesetsky (2005): Cyclic Linearization of Syntactic Structure, *Theoretical Linguistics* 31, 1–45.
Frampton, John & Sam Gutmann (1999): Cyclic Computation, *Syntax* 2, 1–27.
Freidin, Robert (1992): *Foundations of Generative Syntax*. MIT Press, Cambridge, Mass.
Freidin, Robert (1999): Cyclicity and Minimalism. In: S. D. Epstein & N. Hornstein, eds., *Working Minimalism*. MIT Press, Cambridge, Mass, pp. 95–126.
Fukui, Naoki (1997): Attract and the A-over-A Principle, *UCI Working Papers in Linguistics* 3, 51–67.
Gallego, Ángel (2007): Phase Theory and Parametric Variation. PhD thesis, Universitat Autónoma de Barcelona, Barcelona.
Gallego, Ángel (2020): Strong and Weak 'Strict Cyclicity' in Phase Theory. Ms., Universitat Autònoma de Barcelona.
Gazdar, Gerald (1982): Phrase Structure Grammar. In: *The Nature of Syntactic Representation*. Reidel, Dordrecht, pp. 131–186.
Georgi, Doreen (2017): Patterns of Movement Reflexes as the Result of the Order of Merge and Agree, *Linguistic Inquiry* 48(4), 585–626.
Georgi, Doreen & Gereon Müller (2010): Noun Phrase Structure by Reprojection, *Syntax* 13(1), 1–36.
Graf, Thomas (2013): Local and Transderivational Constraints in Syntax and Semantics. PhD thesis, UCLA.
Grimshaw, Jane (1998): Constraints on Constraints in Optimality Theoretic Syntax. Ms., Rutgers University, New Brunswick, New Jersey.
Grohmann, Kleanthes & Andrew Nevins (2004): On the Syntactic Expression of Pejorative Mood, *Linguistic Variation Yearbook* 4, 137–179.
Grohmann, Kleanthes K. (2000): Prolific Peripheries: A Radical View From the Left. PhD thesis, University of Maryland.
Grohmann, Kleanthes K. (2003): Successive Cyclicity under (Anti-)Local Considerations, *Syntax* 6, 260–312.
Haegeman, Liliane (1995): *An Introduction to Government and Binding Theory*. 2 edn,

Blackwell, Oxford.
Hauser, Marc, Noam Chomsky & W. Tecumseh Fitch (2002): The Faculty of Language: What Is It, Who Has It, and How Did It Evolve?, *Science* 298, 1569–1579.
Heck, Fabian (2008): *On Pied-Piping. Wh-Movement and Beyond*. Vol. 98 of *Studies in Generative Grammar*, Mouton de Gruyter, Berlin.
Hornstein, Norbert (2009): *A Theory of Syntax: Minimal Operations and Universal Grammar*. Cambridge University Press, Cambridge.
Hornstein, Norbert, Jairo Nu nes & Kleanthes Grohmann (2005): *Understanding Minimalism*. Cambridge University Press, Cambridge.
Huang, Cheng-Teh James (1982): Logical Relations in Chinese and the Theory of Grammar. PhD thesis, MIT, Cambridge, Mass.
Jacobson, Pauline & Paul Neubauer (1974): Extraposition Rules and the Cycle. In: *Berkeley Studies in Syntax and Semantics*. Vol. 1:viii, pp. 1–99.
Keine, Stefan (2016): Probes and Their Horizons. PhD thesis, University of Massachusetts, Amherst.
Kitahara, Hisatsugu (1997): *Elementary Operations and Optimal Derivations*. MIT Press, Cambridge, Mass.
Koizumi, Masatoshi (1995): Phrase Structure in Minimalist Syntax. PhD thesis, MIT, Cambridge, Mass.
Koster, Jan (1978): *Locality Principles in Syntax*. Foris, Dordrecht.
Kramer, Ruth (2007): The Amharic Definite Marker and the Syntax/PF Interface. Ms., University of California at Santa Cruz.
Kuno, Susumo (1973): Constraints on Internal Clauses and Sentential Subjects, *Linguistic Inquiry* 4, 363–385.
Lakoff, George (1971): On Generative Semantics. In: D. Steinberg & L. Jakobovits, eds., *Semantics*. Cambridge University Press, Cambridge, pp. 232–296.
Lasnik, Howard & Mamoru Saito (1984): On the Nature of Proper Government, *Linguistic Inquiry* 15, 235–289.
Lasnik, Howard & Mamoru Saito (1992): *Move* $\alpha$. MIT Press, Cambridge, Mass.
Lebeaux, David (1988): Language Acquisition and the Form of the Grammar. PhD thesis, University of Massachusetts, Amherst.
Legate, Julie Anne (2005): Phases and Cyclic Agreement, *MITWPL* 49, 147–156. Perspectives on Phases.
Lohndal, Terje (2010): Freezing Effects and Objects, *Journal of Linguistics* 47, 163–199.
Marušič, Franc (2005): On Non-Simultaneous Phases. PhD thesis, Stony Brook University.
Matushansky, Ora (2005): Going Through a Phase, *MITWPL* 49, 157–181. Perspectives on Phases.
Matushansky, Ora (2006): Head-Movement in Linguistic Theory, *Linguistic Inquiry* 37, 69–109.
McCawley, James (1984): Exploitation of the Cyclic Principle as a Research Strategy in Syntax. In: W. de Geest & Y. Putseys, eds., *Sentential Complementation*. Foris, Dordrecht, pp. 165–183.
McCawley, James (1998): *The Syntactic Phenomena of English*. University of Chicago Press, Chicago.
Medeiros, David (2008): Optimal Growth in Phrase Structure, *Biolinguistics* 2, 152–195.
Müller, Gereon (1998): *Incomplete Category Fronting*. Kluwer, Dordrecht.
Müller, Gereon (2004): Argument Encoding and the Order of Elementary Operations. Ms., Universität Leipzig.
Müller, Gereon (2011): *Constraints on Displacement. A Phase-Based Approach*. Vol. 7 of

*Language Faculty and Beyond*, Benjamins, Amsterdam.
Müller, Gereon (2014): *Syntactic Buffers*. Vol. 91, Linguistische Arbeits Berichte, Universität Leipzig.
Müller, Gereon & Wolfgang Sternefeld (2001): The Rise of Competition in Syntax: A Synopsis. In: G. Müller & W. Sternefeld, eds., *Competition in Syntax*. Mouton de Gruyter, Berlin, pp. 1–68.
Perlmutter, David & Scott Soames (1979): *Syntactic Argumentation and the Structure of English*. The University of California Press, Berkeley.
Pesetsky, David (1982): Paths and Categories. PhD thesis, MIT, Cambridge, Mass.
Pesetsky, David (2013): Phrasal Movement and its Discontents: Diseases and Diagnoses. In: L. Cheng & N. Corver, eds., *Diagnostics for Movement*. Oxford University Press, Oxford, pp. 123–157.
Pesetsky, David (2016): Exfoliation: Towards a Derivational Theory of Clause Size. Ms., MIT, Cambridge, Mass.
Pesetsky, David & Esther Torrego (2006): Probes, Goals and Syntactic Categories. Ms., MIT. Lingbuzz/000321.
Polinsky, Maria & Eric Potsdam (2001): Long-Distance Agreement and Topic in Tsez, *Natural Language and Linguistic Theory* 19, 583–646.
Postal, Paul (2012): Two Case Studies of Chomsky's Play Acting at Linguistics. Ms., Lingbuzz.
Preminger, Omer (2018): How to Tell a Syntactic Phenomenon When You See It. Ms., University of Maryland.
Pullum, Geoffrey (1979): *Rule Interaction and the Organization of a Grammar*. Garland, New York.
Pullum, Geoffrey (1992): The Origins of the Cyclic Principle. In: *Proceedings of the Regional Meeting of the Chicago Linguistic Society*. Vol. 28.2, pp. 209–236.
Rackowski, Andrea & Norvin Richards (2005): Phase Edge and Extraction, *Linguistic Inquiry* 36, 565–599.
Reinhart, Tanya (1976): The Syntactic Domain of Anaphora. PhD thesis, MIT, Cambridge, Massachusetts.
Richards, Marc (2007): On Feature Inheritance, *Linguistic Inquiry* 38, 563–572.
Richards, Marc (2011): Deriving the Edge: What's in a Phase?, *Syntax* 14, 74–96.
Richards, Norvin (2001): *Movement in Language*. Oxford University Press, Oxford.
Riemsdijk, Henk van (1978): *A Case Study in Syntactic Markedness: The Binding Nature of Prepositional Phrases*. Foris, Dordrecht.
Riemsdijk, Henk van & Edwin Williams (1986): *Introduction to the Theory of Grammar*. MIT Press, Cambridge, Mass.
Rizzi, Luigi (1982): *Issues in Italian Syntax*. Foris, Dordrecht.
Rizzi, Luigi (1990): *Relativized Minimality*. MIT Press, Cambridge, Mass.
Ross, John (1967): Constraints on Variables in Syntax. PhD thesis, MIT, Cambridge, Mass.
Safir, Ken (2019): The A/A-bar Distinction as an Epiphenomenon, *Linguistic Inquiry* 50, 285–336.
Sauerland, Uli (1999): Erasability and Interpretation, *Syntax* 3, 161–188.
Takahashi, Shoichi & Sarah Hulsey (2009): Wholesale Late Merger: Beyond the A/A-bar Distinction, *Linguistic Inquiry* 40, 387–426.
Takano, Yuji (1994): Unbound Traces and Indeterminacy of Derivation. In: M. Nakamura, ed., *Current Topics in English and Japanese*. Hituzi Syobo, Tokyo, pp. 229–253.
Urk, Coppe van (2015): A Uniform Syntax for Phrasal Movement: A Dinka Bor Case Study. PhD thesis, MIT, Cambridge, Mass.

Vicente, Luis (2007): The Syntax of Heads and Phrases. PhD thesis, LOT: Netherlands Graduate School of Linguistics, Universiteit Leiden.

Wasow, Tom (1972): Anaphoric Relations in English. PhD thesis, MIT, Cambridge, Mass.

Wexler, Ken & Peter Culicover (1980): *Formal Principles of Language Acquisition*. MIT Press, Cambridge, Mass.

Zaenen, Annie & Jessie Pinkham (1976): The Discovery of Another Island, *Linguistic Inquiry* 7, 652–664.