1 Never such innocence again

1.1 A SHOT THAT THANKS TO ROSS WAS HEARD AROUND THE WORLD

It all began so innocently. While discussing issues regarding levels of adequacy in the formulation of grammatical theories, Chomsky 1964 touched on the formulation of transformations such as relative clause formation and question formation, and in so doing proposed what is, to the best of my knowledge, the very first general constraint on long-distance dependencies in natural languages – the very constraint that Ross took as his point of departure in his 1967 thesis, where the notion of island was introduced in linguistic theory.

Chomsky (1964:930–931) writes the following (the original numbering of the examples has been retained):

> Consider the sentences:
>
> (6) (i) Who(m) did Mary see walking toward the railroad station?
> (ii) Do you know the boy who(m) Mary saw walking to the railroad station?
> (7) Mary saw the boy walking toward the railroad station
>
> (7) is multiply ambiguous; in particular it can have either the syntactic analysis (8i) or (8ii)
> (8) (i) NP – Verb – NP – Complement
> (ii) NP – Verb – NP
>
> where the second NP in (8ii) consists of an NP ("the boy") with a restrictive relative clause. The interpretation (8ii) is forced if we add "who was" after "boy" in (7); the interpretation (8i) is forced if we delete "ing" in (7). But (6i, 6ii) are not subject to this ambiguity; the interpretation (8ii) is ruled out, in these cases. Once again, these are facts that a grammar would have to state to achieve descriptive adequacy. . .
>
> The problem of explanatory adequacy is again that of finding a principled basis for the factually correct description. Consider how (6i) and (6ii) must be generated in a transformational grammar of English. Each must be formed by a transformation from a terminal string

> S underlying (7). In each case, a transformation applies to S which selects the second NP, moves it to the front of the string S, and replaces it by a wh-form . . . But in the case of (7) with the structural description (8ii), this specification is ambiguous, since we must determine whether the second NP – the one to be prefixed – is "the boy" or "the boy walking to the railroad station", each of which is an NP. Since transformations must be unambiguous, this matter must be resolved in the general theory. The natural way to resolve it is by a general requirement that the dominating, rather than the dominated, element must always be selected in such a case. This general condition, when appropriately formalized, might then be proposed as a hypothetical linguistic universal. What it asserts is that if the phrase X of category A is embedded with a larger phrase ZXW which is also of category A, then no rule applying to the category A applies to X but only to ZXW.

Let me unpack this important passage a little. Basically, in the paragraphs I have just reproduced, Chomsky points out that the relevant[1] ambiguity that exists in *Mary saw the boy walking to the railroad station* (either Mary saw the boy who was walking to the station or Mary saw the event of walking by the boy) is lost if we relativize or question the sentence as Chomsky does in sentences (8i, 8ii). Chomsky's concern is: Why should this be? The great novelty of Chomsky's (1964) proposal concerns the general nature of the constraint he proposes (cf. "this matter must be resolved in the *general* theory"; "resolve it . . . by a *general* requirement"; "this *general* condition"; "a hypothetical linguistic *universal*"; "*no* rule . . ."): if there is a case where a rule can apply to either a dominating or a dominated element of a given type A, pick the dominating, not the dominated, element. This constraint, which is unnamed in Chomsky (1964) (notice that the constraint is not even presented in indented form, or anything of the sort), but which came to be known as the "A-over-A condition" after Ross referred to it that way in his 1967 thesis, is quite different from something that Chomsky could have suggested: instead of proposing a constraint that covers all movement transformation (cf. "*no* rule . . ."), Chomsky could have tried to incorporate the relevant constraint into the relevant individual transformations he was discussing (question-formation, relative clause-formation, etc.). By formulating the A-over-A condition as a general requirement, Chomsky essentially freed up the individual transformations from the burden of having to incorporate the restriction. As a result, the transformations themselves can be stated in simpler ways.

Ross stresses this very point at the beginning of his thesis (pp. 6–7):

> It is probably unnecessary to point out that it is commonplace to limit the power of the apparatus which is available for the description of particular languages by 'factoring out' of individual grammars

> principles, conditions, conventions and concepts which are necessary in all grammars: to factor out in this manner is to construct a theory of language. So, for example, when the principle of operation of the syntactic transformation cycle has been specified in linguistic theory, it is unnecessary to include another description of this principle in a grammar of French. . . . The present work should be looked upon as an attempt to add to this list a precise specification of the notion *syntactic variable*. The notion is crucial for the theory of syntax, for without it the most striking fact about syntactic processes – the fact that they may operate over indefinitely large domains – cannot be captured. And since almost all transformations either are most generally stated, or can only be stated, with the help of variables, no transformation which contains variables in its structural index will work properly until syntactic theory has provided variables which are neither too powerful nor too weak. It is easy to construct counterexamples . . . for almost every transformation containing variables that has ever been proposed in the literature on generative grammar. It is for this reason that attempts to constrain variables . . . are so important: without the correct set of constraints, it is impossible to formulate almost all syntactic rules precisely, unless one is willing to so greatly increase the power of the descriptive apparatus that every variable in every rule can be constrained individually. But one pursuing this latter course will soon come to realize that many of the constraints he imposes on individual variables must be stated again and again; that he is missing clear generalizations about language. Thus, the latter course must be abandoned: the only possible course is to search for universal constraints.

I think that Ross's words are very clear to the modern reader, except perhaps the notion of 'variable.' Recall that in those early days transformations were formulated *Syntactic Structures*-style: in terms of a structural analysis ("If you find a string such and such . . .") and a structural change ("Turn string such and such into . . .").[2] So, for example, "If you find the following string, 'C_x WYZ X', turn it into 'X-C_x WYZ.'" WYZ were called variables, which provided the context around which the relevant transformation (in our example, joining C_x and X) operated. What Chomsky discovered in 1964 was the need to impose a general constraint on variables in syntax: if you try to move an element of category A, and the context of that transformation is such that this element of category A is dominated by an element of the same category, you must move that bigger, dominating element.

As Ross's remarks make clear, it wasn't the first time that linguists realized that not every property involved in the formulation of a given transformation must be stated in the transformation: there are notions that belong to the general (meta)theory. You don't want to define the notion of phrase as part of the transformation that moves, say, *wh*-phrases.

Likewise, you don't want to have to define the notion of 'verb' every time you want to express that a certain suffix attaches to verbs. Already in the work that provided the context for his own Ph.D. thesis, "The logical structure of linguistic theory" (1955/1975), Chomsky formulated general properties of the theory of grammar. For example, he noted that passive questions (e.g. *Was Bill killed by Sue?*) don't require a special transformation, as they emerge naturally from the interaction of two independently needed transformations (passivization and question-formation). Nevertheless, in that same work, Chomsky missed the opportunity to formulate the very first general locality constraint on long-distance dependencies.[3] On p. 437, Chomsky observes a certain restriction on question-formation by giving the following unacceptable example:

(1) *Whom did [your interest in __] seem to me rather strange?

But unlike what he did in 1964, Chomsky suggested incorporating whatever constraint is operative in (1) into the transformation itself. This suggests that different transformations could be subject to different locality constraints.

The perspective in Chomsky (1964) and in Ross (1967), and for much of subsequent linguistic theorizing,[4] was dramatically different. The focus there was on extracting general conditions and formulating hypothetical linguistic universals – in Ross's terms, quoted above, "to construct a theory of language." Not individual grammars, but universal grammar. I am stressing this because without this theoretical stance, without this universalist aspiration, islands would not be a topic of inquiry (nor would linguistic theory be what it is today). Without this universalist craving, islands would not have been discovered. Islands indeed offer a powerful and enduring illustration of the idea that theories act like microscopes and telescopes; theories are perhaps the most powerful tools for empirical discovery.

There is a final remark I want to make in the context of the passage from Chomsky (1964) quoted above before I turn to later developments in the theory of islands. As I mentioned above, this passage by Chomsky is rightly regarded as the seed that gave birth to Ross's dissertation and all subsequent works on locality, but I think that it is in fact even richer than it looks. It certainly contains the A-over-A constraint, but examining it more closely, it becomes clear that this passage contains a second condition (constraint) on transformations. It is stated even more innocently than the A-over-A condition, as part of "background knowledge," buried in a *since*-clause: "since transformations must be unambiguous, . . ." This too is a general constraint, to be stated "in the general theory," a property of Universal Grammar.

As far as I know, Ross also took this condition for granted, but, as we will see later on in this book (Chapter 4), several linguists subsequently elevated the concern for unambiguity to the level of important principles responsible for island effects (though they did not, as far as I know, point out that the first hint of such principles went back to Chomsky's famous A-over-A passage).

1.2 ISLAND HO!

Be that as it may, Ross did not ignore Chomsky's A-over-A condition. In fact, his entire thesis revolves around it. More precisely, Ross's entire thesis tries to 'fix' the A-over-A condition, for, as Ross pointed out, Chomsky's proposal is both too weak and too strong. It's too weak because there are many examples of illicit question formation about which the A-over-A hypothesis is silent. For instance, nothing seems to ban extraction of the adjective from the Noun Phrase in (2c), even though the moving element is not of the same type/category as the domain it moves from (NP):

(2) a. You have a very nice car
b. How nice a car do you have?
c. *How nice do you have [__ a car]?

The A-over-A is too strong in ruling out acceptable examples of extraction of a Noun contained inside a bigger Noun Phrase, as the following examples show:

(3) a. Who would you approve of [my seeing __]?
b. Which author did you read [a book about __]?

After pointing out that none of the solutions that Chomsky proposed in other versions of his (1964) work were successful, Ross went on to propose more adequate constraints on variables in syntax. Most of these came to take the form "transformational rules of type such and such cannot take place in environment so and so," and 'environment so and so' came to be called an island. Thus, Ross was the first to observe that extraction was not possible out of (among other structural domains) "complex Noun Phrases" (say, an NP modified by a relative clause), "coordinate structures," "sentential subjects," and "left branches" (NPs on the left branches of bigger NPs):

(4) a. *Which book did John meet [a child who read __]? *Complex NP*
b. *What did you eat [ham and __]? *Coordinate Structure*
c. *Who did [that Mary kissed] bother you? *Sentential Subject*
d. *Whose did you buy [__ book]? *Left Branch*

Ross's main concern in his thesis was to characterize as accurately as possible the contexts in which transformations could apply. In so doing, Ross made crucial observations concerning the nature of islands. Let me list those that I think continue to play a significant role in current linguistic theorizing.

First, Ross observed that at least some of these island constraints were language-specific. For example, the Left Branch condition ("No NP which is the leftmost constituent of a larger NP can be extracted out of that NP"), illustrated in (4d), does not appear to hold in many Slavic languages. Likewise, Ross observes that many languages impose a ban on NP-extraction out of a PP ("Preposition-stranding"), but English is not one of them:

(5) Who did you talk [to __]?

Second, Ross noted that languages resort to a variety of strategies to circumvent islands. Thus, many languages, including English, allow for extraction out of a coordinate structure if said extraction takes place "across the board" (in parallel from both conjuncts), as shown in (6).

(6) a. *Which movie did [John hate __ and Bill criticize the book]?
b. Which movie did [John hate __ and Bill criticize __]?

In the same vein, Ross pointed out that many islands can be circumvented if the island is carried along with the element to be moved (a process that Ross made famous by the name of Pied-Piping). This is how English gets around Left Branch Condition violations, and how many languages avoid stranding prepositions.

(7) a. *Whose did you read [__ book]?
b. [Whose book] did you read?

Finally, Ross also pointed out that islands should not be defined in absolute terms. That is to say, Ross showed that it is simply not the case that no rule of any sort is blocked in the presence of an island. Ross's main concern was with a certain class of so-called "reordering transformations"; in a more traditional idiom: with movement processes that leave a gap (so-called "chopping rules"). Other transformations, for example reordering ('movement') rules that leave a pro(nominal) form (a so-called 'resumptive' element) instead of a gap – "copying rules" – appear to be immune to island effects. Witness the difference in behavior in a Complex NP environment in (8):

(8) a. *Who did Sue read [the claim that __ was drunk] in the *Times*?
b. That man, Sue read [the claim that *he* was drunk] in the *Times*

In fact, these are Ross's last words on this matter in his Ph.D. thesis (right before the concluding chapter):

> Variables in chopping rules, feature-changing rules, and unidirectional rules of deletion cannot cross island boundaries; variables in other rules can (p. 289)

In those days, it was standard to distinguish among various kinds of rules, and Ross certainly did so. In addition to deletion rules ('ellipsis'), he appealed to feature-changing rules (e.g., today's Negative Polarity Item licensing operation, which in those days was a rule turning *some* into *any*), and re-ordering rules, and in the context of the latter, he distinguished among those reordering rules which left a pro(nominal) form (copying rules), and those that left a gap (chopping rules). He furthermore distinguished between chopping rules that were leftward-oriented and those that were rightward-oriented because only leftward chopping rules allowed for Preposition stranding in English, and were not upward bounded as rightward chopping rules were (rightward chopping rules were restricted to apply to their own clause; i.e., they were "clause-bounded," which Ross dubbed the Right-Roof Constraint).

I am stressing this fact because as we are about to see, subsequent theorizing on islands, probably under the influence of Chomsky (1973, 1977), has tended to view islands as domains out of which any form of movement was prohibited, and has treated those dependencies crossing islands in non-movement terms (so-called "rules of construal"). For this reason, it is common to come across definitions of islands, in textbooks and elsewhere, as the following:[5]

> We say that a phrase is an "island" if it is immune to the application of rules that relate its parts to a position outside of the island. Thus to say that a *wh*-clause is an island is to say, in particular, that the rule of *wh*-movement that forms questions and relatives by moving such expressions as *who*, *what, what sonatas*, etc., to the left of a clause cannot be applied in general to a *wh*-expression with a *wh-clause*. (Chomsky 1980:194)

> "Islands" are syntactic configurations . . . into which the relation of *wh* binding may not reach. (McCloskey 1988:23)

> islands . . . Syntactic configurations which do not permit movement rules . . . to move categories from positions inside them to positions outside them. (Roberts 1997:284)

> "Islands" is the cover term for nodes which obstruct syntactic movement (Szabolcsi and den Dikken 2002:213)

A quote from an anonymous reviewer reported in Postal (1997:2), sums up the standard take on islands rather well: "If something allows extraction, then it is not an island. This is at least the current view of the notion island. The copying rules of Ross (1967) today would not be taken to involve extraction, at least not when they are not island-sensitive."

The following quote, from Freidin (1992:94),

> A construction from which a constituent may not be moved by a transformation is designated as an island (following Ross 1967)

illustrates that this standard position is often attributed to Ross, but this is inaccurate. Ross's position on this point was far more nuanced, far more flexible, and, as we will see as we proceed, quite likely also much closer to the truth. (I suspect that Ross's more qualified stance on islands stems from the starting point of his thesis: the recognition that Chomsky's A-over-A hypothesis turned out to be too strong.)

What is accurate about Ross's position, and – as Postal (1997) insightfully remarked – not obviously true, is that he took syntactic domains to be non-islands by default. That is, as Ross stresses throughout his thesis, he took as a fundamental, basic property of human language (indeed, as Ross says on p. 7, "the most striking fact about syntactic processes") that syntactic dependencies were unbounded; they may operate over indefinitely large domain. Alongside Postal, I want to point out that this is a reasonable, but by no means obvious, assumption, certainly from a modern viewpoint where the family of domains that count as islands has grown a lot since Ross's early proposals. Perhaps the domains that allow for chopping rules that leave a gap are the exception rather than the rule.

Be that as it may, looking back at Ross's study, syntacticians like myself feel very fortunate indeed that Ross did not ignore Chomsky's brief discussion of the A-over-A condition, for it led to a reorientation of linguistic theory. As Postal (1986:xvii) writes in his Foreword to *Infinite Syntax!* (the long-awaited book-version of Ross's thesis),

> Previously, attempts to construct fragments of transformational grammars had overwhelmingly tended to assume that restrictions on particular constructions relevant to a hypothesized rule had to be built into the structure of that rule. In practice, this led to postulated rules of extraordinary complexity, involving myriads of *ad hoc* constraints. It further led to a lack of comparability between rules for different constructions, and still more for different languages. It obscured the possibility that large classes of different constructions were subject to similar constraints.

In sum, Ross's study, by "concretiz[ing] the abstract possibility of general constraints on rules in a set of actual proposals [that], while hardly perfect or the last word on the matter, were sufficiently close to the mark to have continued to be the basis for further work through the present day" (Postal 1986:xviii–xix), planted the seeds of a theory of Universal Grammar, where constructions in specific languages are epiphenomena, and the seeds of the new, revitalized, Comparative Syntax of today, where constructions can be compared across languages, and in doing so, Ross pointed to a very fruitful way of addressing "Plato's problem," the logical problem of language acquisition.

1.3 FIRST DREAMS OF A FINAL THEORY

The first author to fulfill the potential of Ross's proposals was Chomsky who, in his (1973) article "Conditions on transformations," took the goal of extracting general conditions from specific constructions to new heights, as he set out to uncover what the various island configurations discovered by Ross had in common. In fact, Chomsky did more than this. His (1973) article is an attempt to uncover and unify all the locality principles constraining transformations. Thus, Chomsky does not begin "Conditions on transformations" with island effects, but with more stringent locality conditions, such as the impossibility of relating an anaphor to its antecedent across a tensed clause or across an overt, lexical (there called "specified") subject:

(9) a. *John said [that himself was smart]
b. *John said [that Sue liked himself]

Based on the observation that the same constraint appears to hold of movement (10), Chomsky proposed the following conditions (11)/(12), which he named the "Tensed-S Condition" and the "Specified Subject Condition," respectively.

(10) a. *John seems [that __ is smart]
b. *John seems [that Mary likes __]

(11) No rule can involve *X*, *Y* in the structure
... *X*... [$_{\alpha}$ ... *Y* ...] ...
where α is a tensed sentence.

(12) No rule can involve *X*, *Y* in the structure
... *X*... [$_{\alpha}$ ... *Z* ... –WYV ...] ...
where *Z* is the specified subject of *WYV* in α.

Notice the universalist character of the rule "no rule . . ."; no rule . . . of any kind (recall, and contrast with, Ross's careful distinctions among rules that can or cannot cross islands). The rigidity of the conditions, however, turned out to be too strong. As Chomsky realized, some dependencies can be formed across tensed clauses and specified subjects:

(13) a. Who do you believe [__ likes Mary]?
b. Who do you believe [Mary likes __]?

Accordingly, Chomsky had to qualify his original conditions to allow for those instances of movement. Chomsky noticed that the dependencies in questions involved elements that typically occupy complementizer positions (today's CP area) and, moreover, that the conditions could only be violated if the movement really targets the complementizer domain and if there is an unoccupied complementizer position at the edge of the domain out of which movement takes place. The first observation accounts for the fact that it is possible to move a *wh*-word as was done in (13) but impossible to have the *wh*-word land in the subject (non-complementizer) position of the higher clause:[6]

(14) *Who is said that [Mary likes __]?

The second observation captures the fact that the type of movement allowed in (13) is ruled out in the presence of a filled complementizer domain of the relevant Tensed/Specified Subject extraction site:

(15) *Who did you ask [where John saw __]?

The outcome of these observations led to the following revised locality condition:

(16) No rule can involve *X*, *Y* in the structure
. . . *X*. . . [$_{\alpha}$ . . . *Z* . . . –WYV . . .] . . .
where (a) *Z* is the specified subject of *WYV*
or (b) *Y* is in COMP and *X* is not in COMP
or (c) *Y* is not in COMP and α is a tensed S.[7]

From this perspective, Chomsky suggested that the reason that the movement in (13) can violate his original conditions is due to the fact that the moving element can first land in the intermediate COMP position, at the edge of the locality domain characterized by his original conditions, and in so doing circumvent their effects. Thus was born the idea that long-distance movement proceeds in small steps, COMP-to-COMP, or, as it soon came to be known, successive cyclically.

Chomsky then observed that the COMP position that movement exploits to reach the final landing site (the other COMP) should not