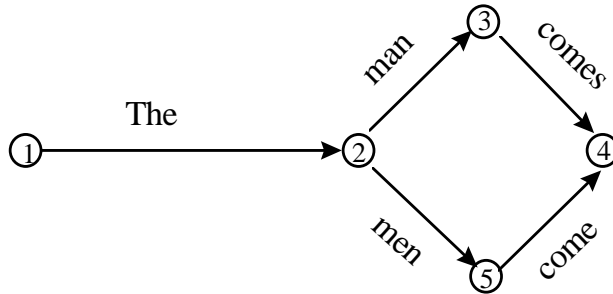# Up and Down the Chomsky Hierarchy

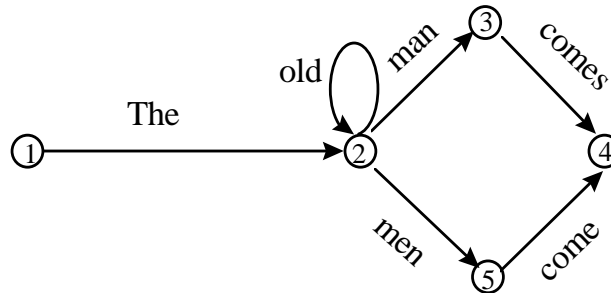Howard Lasnik
University of Maryland
lasnik@umd.edu

Markovian and Non-Markovian Properties in Lower Level Syntax

(1)   Some finite state languages:

(2)   The old man comes / The old men come

(3)

(4)   The man comes / The old man comes / the old old man comes / ...

(5)

(6)   Some non finite state context free languages:

(7)   *ab*, *aabb*, *aaabbb*, ..., and in general, all sentences consisting of n occurrences of *a*
        followed by n occurrences of *b* and only these

(8)   Σ: S
        F: S → aSb
            S → ab

(9)   *aa*, *bb*, *abba*, *baab*, *aaaa*, *bbbb*, *aabbaa*, *abbbba*, ..., and in general, all sentences
        consisting of a string *X* followed by the 'mirror image' of *X* (i.e., *X* in reverse), and only
        these

(10) Σ: S
        F: S $\rightarrow$ aSa
            S $\rightarrow$ bSb
            S $\rightarrow$ aa
            S $\rightarrow$ bb

(11)   Chomsky argues that (portions of) English cannot be described by finite state grammars:

(12)a   If S_1, then S_2
    b   Either S_3, or S_4
    c   The man who said that S_5, is arriving today

As Chomsky observes, the crucial property of these examples is not merely that there can be a
string of unlimited length between the dependent items (if-then, either-or, man-is). There can
also be a string of unlimited length between *the* and *man* in the finite state language (5). Rather,
there is recursion between the dependent items.

> "In [(12)a], we cannot have "or" in place of "then"; in [(12)b], we cannot have "then" in
> place of "or"; in [(12)c], we cannot have "are" instead of "is". In each of these cases there
> is a dependency between words on opposite sides of the comma (i.e., "if"-"then",
> "either"-"or", "man"-"is"). But between the interdependent words, in each case, we can
> insert a declarative sentence S_1, S_3, S_5, and this declarative sentence may in fact be one of
> [(12)a-c] ... It is clear, then, that in English we can find a sequence $a + S_1 + b$, where
> there is a dependency between *a* and *b*, and we can select as S_1 another sequence
> containing $c+S_2+d$, where there is a dependency between *c* and *d*, then select as S_2
> another sequence of this form, etc. A set of sentences that is constructed in this way ...
> will have all of the mirror image properties of [(9)] which exclude [(9)] from the set of
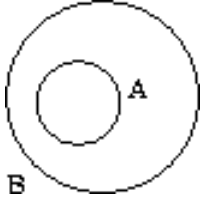> finite state languages."   p.22

(13)   anti^n missile^{n+1} (n *anti*'s followed by n+1 *missile*'s)   [ex. due to Morris Halle,
                                        MIT class lecture 1968]

(14) Σ: W
        F: W $\rightarrow$ anti W missile
            W $\rightarrow$ missile

Σ, F grammars are wonderfully capable of handling languages with the properties of those in (7)
and (9). Further, they can easily generate all finite state languages as well.
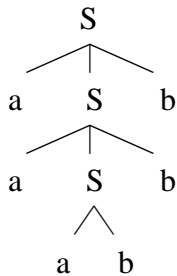
(15)   "A set of strings is called a terminal language if it is the set of terminal strings for some
        grammar [Σ, F]. Thus each such grammar defines some terminal language (perhaps the
        'empty' language containing no sentences), and each terminal language is produced by
        some grammar of the form [Σ, F]."   p.30

(16)   Theorem: Every finite state language is a terminal language, but there are terminal languages which are not finite state language.   p.30

A weak generation subset relation:

(17)

A = finite state (E) languages (weakly generated)
B = context-free (E) languages (weakly generated)

(18)   The fundamental descriptive advantage of PS grammars compared to finite state grammars is that PS grammars can pair up things that are indefinitely far apart, and the way they do it is by introducing symbols that are never physically manifested: the non-terminal symbols. That is, they automatically introduce structure. I return to this issue.

(19)

There is actually a certain irony that arises at this point. While languages abstractly like (7) and (9) were presented as motivating the move from finite state description to Σ, F description, Chomsky did not actually use the crucial relevant properties of Σ, F grammars in his linguistic work at the time.

(20)   The relevant property is, of course, unbounded self-embedding. However, the theory of Chomsky (1955), assumed in Chomsky (1957), has no recursion in the base.

(21)   Instead, it is the transformational component that accounts for the infinitude of language. This point is only hinted at in Chomsky (1957, p.80), but is fully developed in Chomsky (1955, pp. 518 and 526).

(22)   Further, while Σ, F description certainly can weakly generate the English fragments in question, it isn't clear that they are all correctly strongly generated.

(23)   Consider again (12)a: If S_1, then S_2

(24)   "... there is a dependency between words on opposite sides of the comma (i.e., 'if' - 'then' ...)."   p.22

(25)   "It is clear, then, that in English we can find a sequence $a + S_1 + b$, where there is a dependency between *a* and *b*, and we can select as S_1 another sequence containing $c+S_2+d$, where there is a dependency between *c* and *d*, then select as S_2 another sequence of this form, etc. A set of sentences that is constructed in this way ... will have all of the mirror image properties of [(9)] which exclude [(9)] from the set of finite state languages."   p.22

(26)   S $\rightarrow$ if S then S

This captures the unbounded dependency between *if* and *then*. But is it correct in terms of strong generation? Not if *if* S is a constituent, as is most likely the case.
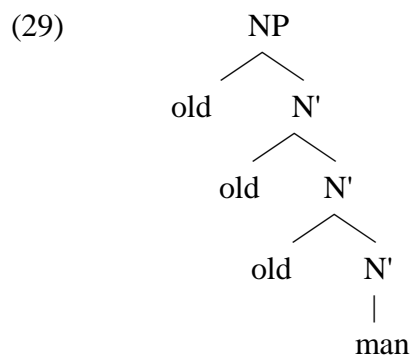
One of the major benefits of Σ, F description is that, unlike finite state description, it automatically and unavoidably provides sentences with structure. This is overwhelmingly positive since, alongside infinitude, constituent structure is the most fundamental and universal property of human languages.

BUT there are rare exceptions, as discussed by Chomsky (1961, p.15) and Chomsky and Miller (1963, p.298). One of the most striking ones is what Chomsky called "true coordination" as in (4), repeated here.

(27)   The man comes / The old man comes / the old old man comes / ...

(28)   "Immediate constituent analysis has been sharply and, I think, correctly criticized as in general imposing too much structure on sentences."

There is no evident syntactic, semantic, or phonological motivation for a structure in which, say, each old modifies the remaining sequence of olds plus man, as in (29), or some such (with irrelevant details omitted).

(29)
```
                NP
              /    \
          old       N'
                  /    \
              old       N'
                      /    \
                  old       N'
                            |
                           man
```

Preferable might be something like

(30)

```
                 NP
               /    \
          Adj.        N'
         /  |  \       |
      old  old old   man
```
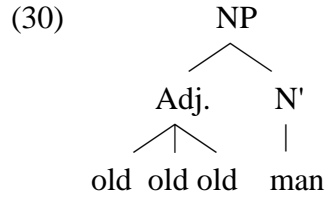
Chomsky mentions a similar case:

(31)   the man was old, tired, tall ..., but friendly

(32)   "The only correct P-marker would assign no internal structure at all within the sequence
        of coordinated items. But a constituent structure grammar can accommodate this
        possibility only with an infinite number of rules; that is, it must necessarily impose
        further structure, in quite an arbitrary way." [p. 15]

Chomsky and Miller (1963, p.298) present a very similar argument:

(33)   "... a constituent-structure grammar necessarily imposes too rich an analysis on sentences
        because of features inherent in the way *P*-markers are defined for such sentences."

With respect to an example identical to (31) in all relevant respects, they say:

(34)   "In order to generate such strings, a constituent-structure grammar must either impose
        some arbitrary structure (e.g., using a right recursive rule), in which case an incorrect
        structural description is generated, or it must contain an infinite number of rules. Clearly,
        in the case of true coordination, by the very meaning of the term, no internal structure
        should be assigned at all within the sequence of coordinate items."
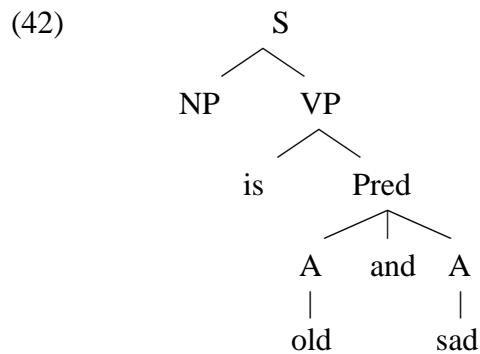
(35)   The conclusion of Chomsky and of Chomsky and Miller: We need to go beyond the
        power of Σ, F description to adequately describe natural languages. In particular, the
        model is augmented by a transformational component.

Chomsky (1955) had, of course, already shown in great detail how transformations can provide
natural accounts of phenomena that can only be described in cumbersome and unrevealing ways
(if at all) by Σ, F grammars. But he had little to say then about the "too much structure" problem
we are now considering.

Chomsky (1961) and Chomsky and Miller (1963) don't have a lot to say either, beyond the
implication that transformations will solve the problem. That is, we need to move up the power
hierarchy.

(36)   In fact, as already mentioned, Chomsky (1955) had already argued that there is no
        recursion in the Σ, F component, the transformational component (in particular
        generalized transformations (GTs)) being responsible in toto for infinitude.

(37)   Chomsky discusses several aspects of the coordination process, though without actually giving a precise formulation of the relevant transformation(s).

(38)   All examples discussed in Chomsky (1955) involve coordination of two items, as in (39).

(39)   John was sad and tired

(40)   For such cases, it is relatively straightforward to formulate an appropriate generalized transformation, even if, as claimed by Chomsky (1961, p.134), GTs are strictly binary (an idea strongly advocated by Kayne (1984)and Kayne (1994) and that is important in Chomsky's recent work.

(41)   As an example, he gives "John is old and sad" from "John is old" "John is sad", with resulting structure (42).

(42)
```
                    S
                  /   \
              NP       VP
                      /  \
                   is     Pred
                         /  \
                      A   and   A
                      |         |
                     old       sad
```

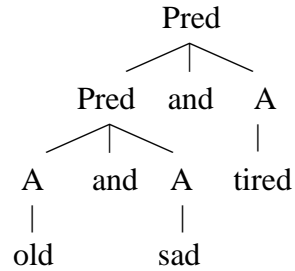Chomsky and Miller also seem to assume binarity, at least in one place in their discussion.

(43)   "The basic recursive devices in the grammar are the generalized transformations that produce a string from a <u>pair</u> [emphasis mine] of underlying strings." [p. 304]

It is not entirely clear what is supposed to happen when we have multiple items coordinated, as in the phenomena principally under discussion here, or in, e.g.:
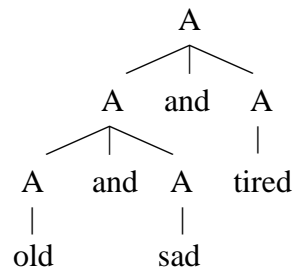
(44)   old and sad and tired

One possibility is that we would preserve the structure of "old and sad" in (42), and create a higher structure incorporating "and tired".
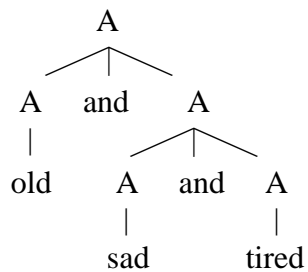
(45)
```
                        Pred
                         /\
                 Pred   and   A
                  /\           |
              A   and   A    tired
              |         |
             old       sad
```

Or, somewhat revising (42):

(46)
```
                         A
                         /\
                  A     and    A
                 /\             |
             A   and   A      tired
             |         |
            old       sad
```

Another possibility is a right branching analogue:

(47)
```
                    A
                    /\
            A    and     A
            |           /\
           old      A   and   A
                    |         |
                   sad      tired
```

BUT any of these would run afoul of Chomsky's argument: In general, we do not want that extra structure.

Yet another possibility, one that would yield the desired 'flatness', arises if we relax the binarity requirement. Chomsky and Miller seemingly countenance this possibility in at least one place in their discussion:

(48)   "We now add to the grammar a set of operations called *grammatical transformations*, each of which maps an <u>*n*-tuple</u> [emphasis mine] of *P*-markers ($n \geq 1$) into a new *P*-marker." [p. 299]

Then a GT could be formulated to coordinate three items (alongside the GT coordinating two items).

BUT of course there is no limit on the number of items that can be coordinated - Chomsky's

original point. So this solution merely replaces one untenable situation with another: In place of an infinite number of phrase structure rules, we have an infinite number of transformations.

(49)   Thus, moving up the power hierarchy ultimately does not help in this instance.
(50)   In a manner of speaking, what we really want to do is move <u>down</u> the hierarchy. Markov processes give flat objects, as they impose no structure.

BUT that is not quite the answer either. While it would work fine for coordination of terminal symbols, phrases can also be coordinated, and, again, with no upper bound. Alongside (51), we find (52).

(51)   John and Bill and Fred and ...
(52)   The old man and the young man and the boy and ...

We need a sort of higher order flatness.

Chomsky and Miller [p. 298] consider, but reject, an extension of constituent structure grammar to yield such flatness. Their extension is, as far as I can tell, equivalent to the so-called Kleene-* device of Kleene (1956). The specific instance they give is:

(53)   Predicate $\rightarrow$ Adj^n and Adj $(n \geq 1)$

Chomsky and Miller indicate that there are "many difficulties involved in formulating this notion so that descriptive adequacy may be maintained ..." But they do not elaborate on this point. It might be interesting to explore this further, since the phenomenon at issue is not exotic and ought to be incorporated into the core theory.

Appendix: Markovian and Non-Markovian Properties in Higher Level Syntax

A. Trace deletion: Subjacency vs. ECP   Chomsky (1991), Chomsky and Lasnik (1993)

(54)   ??Who do you wonder [$_{CP}$ whether [$_{IP}$ John said [$_{CP}$ t' [$_{IP}$ t solved the problem]]]]
                                                          (*)

(55)   Deletion is possible only to turn an illegitimate LF object into a legitimate one, where the
          legitimate LF objects are:
(56)a   Uniform chains (all of whose members are in A-positions; A'-positions; or X^0-positions)
     b   Operator-variable pairs.

(57)   Deletion in the chain (<u>Who</u>, <u>t</u>', <u>t</u>) is permissible since the chain is neither uniform (<u>Who</u>
          and <u>t</u>' are in A'-positions, <u>t</u> in an A-position) nor is it an operator-variable <u>pair</u>.

(58)   More generally, in the case of successive-cyclic A'-movement of an argument, an
          intermediate trace (starred or otherwise) can (in fact must) be deleted in LF, voiding an
          ECP violation when the trace to be deleted is starred.

(59)   On the other hand, long movement as in (60) **will** be an ECP violation, since the
          movement chain in this instance is uniformly A', so economy prevents the deletion of <u>t</u>':

(60)   *How do you wonder [$_{CP}$ whether [$_{IP}$ John said [$_{CP}$ t' [$_{IP}$ Mary solved the problem <u>t</u>]]]]
                                                          (*)

Not a great analysis, perhaps, but it has one advantage over many existing alternatives: It gets
the facts right. The interesting problem that it raises is that inspection of both the ultimate
representation and specific points in the transformational derivation is necessary to correctly
determine the status of the examples. This is in seeming violation of the usually assumed strictly
Markovian nature of transformational derivations.

B. Island violation amelioration

This situation in A. is strikingly reminiscent of one first discussed by Ross (1969):

(61)   I believe that he bit someone, but they don't know who (I believe that he bit)
(62)a *I believe the claim that he bit someone, but they don't know who I believe the claim that
          he bit [Complex NP Constraint, noun complement]
     b(??)I believe the claim that he bit someone, but they don't know who
(63)a *Irv and someone were dancing together, but I don't know who Irv and were dancing
          together [Coordinate Structure Constraint]
     b(??) Irv and someone were dancing together, but I don't know who

(64)a *She kissed a man who bit one of my friends, but Tom doesn't realize which one of my
       friends she kissed a man who bit   [Complex NP Constraint, relative clause]
    b(??)She kissed a man who bit one of my friends, but Tom doesn't realize which one of my
       friends
(65)a *That he'll hire someone is possible, but I won't divulge who that he'll hire is possible
       [Sentential Subject Constraint]
    b (??)That he'll hire someone is possible, but I won't divulge who

All above from Ross (1969), and the judgments are his.

Ross argues that the phenomenon of island violation repair provides "evidence of the strongest
sort that the theoretical power of [global] derivational constraints is needed in linguistic
theory..." [p.277] [that is, that the derivation is non-Markovian].

(66)   If a node is moved out of its island, an ungrammatical sentence will result. If the island-
        forming node does not appear in surface structure, violations of lesser severity will (in
        general) ensue. [p.277]

(67)   Chomsky rejects this kind of global derivational constraint, and suggests [see also Baker
        and Brame (1972), and, for an opposing view, Lakoff (1970), Lakoff (1972)] that * (# in
        Chomsky's presentation) is assigned to an island when it is crossed by a movement
        operation (the complex NP in (70)). An output condition forbidding * in surface
        structures accounts for the deviance of standard island violations.
(68)   If a later operation (Sluicing in this case) deletes a category containing the *-marked
        item, the derivation is salvaged.

(69)a (*)I don't know which children he has plans to send to college
    b   He has plans to send some of his children to college, but I don't know which ones

(70)   I don't know CP
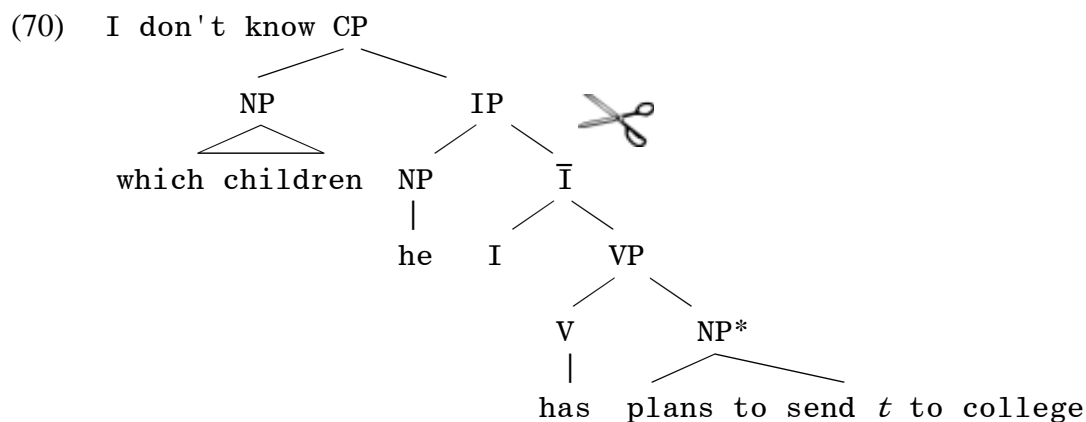
As Lakoff (1972) points out, though, on Ross's judgments substantial globality still remains (as
in the Subjacency vs. ECP phenomenon). For Ross, the examples are improved, but still deviant.
Neither the step in the derivation causing the violation nor the ultimate structure suffices to
determine the status of the examples. Both must be inspected.

Chomsky (1972) briefly addresses this point in his footnote (13), saying "What is at issue is only whether in determining (degree of) deviance, we refer to presence of # only in the surface structure or, alternatively, to its presence anywhere in the derivation as a feature of the category X that is deleted. Thus the issue is one that belongs to the theory of interpretation of deviant structures."

I conclude by noting that potential difficulty both of the phenomena in this section raise for the version of multiple spell out Chomsky has advocated in the last several years (see, for example, Chomsky (2001)) whereby all interpretation is strictly cyclic.

References

Baker, C. L. and Michael Brame. 1972. 'Global rules': A rejoinder. *Language* 48: 51-75.

Chomsky, Noam. 1955. The logical structure of linguistic theory. Ms. Harvard University, Cambridge, Mass. and MIT, Cambridge, Mass.*, .*[Revised 1956 version published in part by Plenum, New York, 1975; University of Chicago Press, Chicago, 1985].

Chomsky, Noam. 1957. *Syntactic structures*. The Hague: Mouton.

Chomsky, Noam. 1961. On the notion "rule of grammar". In *Structure of language and its mathematical aspects*, ed. Roman Jakobson, 6-24. Providence, Rhode Island: American Mathematical Society. [Reprinted in The structure of language: Readings in the philosophy of language, eds. Jerry A. Fodor and Jerrold J. Katz, 119-136. Englewood Clifs, N. J.: Prentice-Hall, 1964].

Chomsky, Noam. 1972. Some empirical issues in the theory of transformational grammar. In *Goals of linguistic theory*, ed. Paul Stanley Peters, 63-130. Englewood Cliffs: Prentice-Hall Inc.

Chomsky, Noam. 1991. Some notes on economy of derivation and representation. In *Principles and parameters in comparative grammar*, ed. Robert Freidin, 417-454. Cambridge, Mass.: MIT Press. [Reprinted in Noam Chomsky, *The minimalist program*, 129-166. Cambridge, Mass.: MIT Press, 1995].

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

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. [Reprinted in Noam Chomsky, *The minimalist program*, 13-127. Cambridge, Mass.: MIT Press, 1995].

Chomsky, Noam and George Miller. 1963. Introduction to the formal analysis of natural languages. In *Handbook of Mathematical Psychology 2*, ed. Duncan R. Luce, Robert R. Bush, and Eugene Galanter, 269-321. New York: Wiley and Sons.

Kayne, Richard. 1984. *Connectedness and binary branching*. Dordrecht: Foris.

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

Kleene, Stephen. 1956. Representation of events in nerve nets and finite automata. In *Automata Studies*, ed. Claude Shannon and John McCarthy, 3-41. Princeton: Princeton University

Press.

Lakoff, George. 1970. Global rules. *Language* 46: 627-639.

Lakoff, George. 1972. The arbitrary basis of transformational grammar. *Language* 48: 76-87.

Ross, John Robert. 1969. Guess who? In *Papers from the Fifth Regional Meeting of the Chicago Linguistic Society*, ed. Robert I. Binnick, Alice Davison, Georgia M. Green, and Jerry L. Morgan, 252-286. Chicago Linguistic Society, University of Chicago, Chicago, Ill.