Summary, notes and extensions of what we saw last week.

-Chomsky, Halle, Lukoff (CHL) 1956 introduced cyclic application of rules in phonology. The term 'cycle' is not used. And, in fact, there is no explicit argument for the cycle. Rather, the arguments are that words have constituent structure and this structure is crucial in determining word stress (including compound words). They argue that to get the stress right, we need particular constituent structures. Then, lo and behold, the needed structures correspond to those determined by syntax and morphology.
They use the constituent structure and cyclic application in their derivations but don't really emphasize it, and don't even articulate it until p.75 top, 2/3 of the way through the paper:
> "Rule 4: Given a phonemic clause,
> (i) assign the value 1 to all accented vowels;
> (ii) then apply each rule pertaining to accented vowels no more than once to each constituent, applying a rule to a constituent of order n only after having applied it to all constituents of order $n + l$; i.e. beginning with the smallest constituents and proceeding to larger and larger constituents;"

There is no further mention of cyclic application, though it continues to be crucially used throughout the paper. There is further mention of constituent structure:
> "A constituent hierarchy has always been considered a characteristic feature of the higher levels of morphology and syntax. We are suggesting here that it exists on the phonological level as well."   p.78 bottom

<In fn12, they attribute to Hockett's <u>Manual of Phonology</u> the claim that phonological representations have nested constituent structure, quoting him: "... the phonologic structure of an utterance shows a hierarchic organization ....">
Given the parallelism they draw with syntax, it becomes even more striking that cyclic application of syntactic rules was several years later. <Speculating here, perhaps this was because GTs, and ban on recursion in the base, were so embedded in LSLT, the root of all generative syntactic work at the time.>

-Halle and Chomsky 1960 are a bit more explicit. On p.276, the 2nd page of the paper, they say that each morpheme in isolation has its own stress pattern, but that this can be altered, depending on the constituent structure of the utterance in which the morpheme is found. About these modifications, they say:
> The modifications are introduced in a stepwise fashion, successive steps reflecting the influence of successively higher constituents. Note also that the same modifications apply to all constituents regardless of their place in the constituent hierarchy; the same rules are reapplied to each constituent in a repeating cycle until the highest constituent is reached. The final result of such a cyclical reapplication of the same rules reflects to a certain extent the stress distribution of the morphemes as parts of lower constituents.

And note that here they do use the terms 'cycle' and 'cyclical reapplication'.

-Fillmore's 1963 paper was the next thing we looked at.

<Chomsky attributes 2 big arguments to it, but they sure aren't explicit (maybe not even implicit). Aspects p. 133 "In particular, there are no known cases of ordering among generalized embedding transformations although such ordering is permitted by the theory of Transformation-markers. Furthermore, there are no really convincing cases of singulary transformations that must apply to a matrix sentence before a sentence transform is embedded in it, though this too is a possibility, according to the theory." <<The T-marker Chomsky displays is, as he notes, consistent with these hypothesized constraints.>>

In fn 5, Chomsky attributes both of these observations to Fillmore. Did F. actually say either of these things?

Chomsky 1966, Topics, is slightly more reserved in his Fillmore attribution. pp.61-62:

"C. J. Fillmore pointed out that there are many restrictions on the organization of T-markers beyond those that were assumed in earlier attempts to formulate a theory of transformational grammar [Fillmore, "The position of embedding transformations in a grammar", Word 19.208-31 (1963)]. What his observations come to is essentially this: there is no ordering among generalized transformations, although singulary transformations are ordered (apparently linearly); there are no singulary transformations that must apply to a matrix sentence before a constituent sentence is embedded in it by a generalized embedding transformation, although there are many instances of singulary transformations that must apply to a matrix sentence after embedding of a constituent structure within it and to a constituent sentence before it is embedded; embedding should be regarded as substitution of a sentence transform for a 'dummy symbol' rather than as insertion of this transform in a categorially unspecified position."

What did Fillmore actually say?
p.208

"To date, the position of these two types of transformations [STs and GTs] with respect to each other has been left unspecified in the published writings of transformational grammarians, and it has been suggested by Lees that there may need to be a complex system of "traffic rules" for indicating the possible ways in which the result of a particular generalized transformation may be channeled back to a specific position in the sequence of simple transformations.

In this paper a general solution to the "traffic rule" problem is proposed, "

Fillmore than gives a diagrammatic presentation of how the kinds of Ts inter-relate. (Fig. 1 on p.209 of his paper), and says "... the result of an embedding transformation passes to the preliminary simple transformation (PST) component ..."

On p. 213, he says

"it appears that it should be possible to produce a terminal string and let it pass through the PST component, producing thereby a pre-sentence. It must then be possible (for some pre-sentences) to embed this pre-sentence into another terminal string, after which the resulting form must now be subject to the PSTs. It should even be possible to choose the same PSTs each time through.

It is possible, in fact, to construct sentences which illustrate a re-cycling through the PSTs in which the same transformation is chosen both times."

He then presents a derivation of a sentence illustrating this interaction on pp.213-215. The sentence is

The butler is believed to have been murdered

The closest thing I can find to one of the arguments Chomsky attributes to Fillmore comes on p.216:   <<There is nothing remotely like the argument that we don't need ordering among GTs.>>

> "... a complex construction, such as a passive, may be embedded into a simple construction, namely a terminal string, and further complexities are interpreted as involving operations on the resulting whole. If terminal strings were to be embedded into terminal strings, then the PSTs would have to be made much more complex in order to allow for the construction of complex embedded sentences, such as embedded passives. If, on the other hand, pre-sentences were to be embedded into pre-sentences, then the embedding transformations would have to become much more complex, in order to allow embedding into any type of sentences, i.e., not only into simple actives, but into passives, questions, embedded passives, and so on."

<<Terminal strings are the structured objects generated by the phrase structure grammar. PSTs are singulary transformations.>>
All of the rest of the paper is devoted to enumerating transformations and giving derivations of sentences. There is no other theoretical or conceptual discussion until the very end of the body of the text, where he says: p.224

> ".In this paper a general solution to the problem of constructing the "traffic rules" in a transformational grammar has been proposed. This solution is believed to lie in the construction of a new representational level, the level of pre-sentences ..."

<<" Pre-sentences are representations which are ready to undergo a final group of obligatory and stylistic rules which yield, as the final output of the grammar, phonetic descriptions of sentences." p.212>>

So let's take a quick look at the Lees paper.
Lees pp.53-54 observes that strict ordering of PS rules is generally automatically determined (what might be called intrinsic ordering). This is not generally so for transformations. Note that when he says

> "we have been forced to build. into the formal features of transformation rules a mechanism capable of taking account also of the entire derivational (and perhaps also transformational) history of a string as input, this limitation of strict ordering is no longer automatically imposed within the transformational part of the grammar."

he is not saying transformational derivations are non-Markovian. Rather, he is saying they are structure dependent. They care about the entire structure (thus, the entire PS derivation), not just the terminal string. And that entire structure can be a derived structure, not just a basic structure. The most relevant passage in the Lees paper is on pp.54-55, but I find it really opaque. Lees starts by observing that the NPs manipulated by Passive can be derived nominals (hence, NPs inserted into a structure by GTs). He indicates that rule ordering of Ts can help reduce the complexity of some Ts. Basically, it doesn't matter what the source of the NP is; Passive won't care. And, I think he is saying, we get this result if the relevant GTs can apply before Passive: p.54

> "For example, in specifying the type of sentence which may undergo the passive

transformation (T2) (p. 34), it is not necessary to make special reference to intervening nominalizations which produce new, derived nominals as subjects of argument sentences, provided only that the "traffic rules" of the grammar permit the application of these nominalizations before the application of the passive in the derivation of any sentence."

But, he notes a difficulty:

" ... these nominalization transformations themselves must be permitted to apply to sentences which have already undergone the passive transformation!"

He gives as an ex.:

We were displeased by his having been arrested too soon.

Apparently he is worried about how we can get a derivation like Passive>GT>Passive if we have strict linear ordering.   <We will talk more about such derivations.>

He then (I think) gives 4 possible solutions to the problem. I don't fully understand any of them, but here we go.

1. <Maybe this one is hinting at the argument that Chomsky gives: We don't need derivations where a rule must apply in the matrix before a constituent sentence is embedded in it.>

" suppose that the first, or new-constituent-forming, part of the transformation occurs in proper order, and then the generated transform, say the nominal, is, so to say, "held in storage, ... Each time thereafter a string fitting the qualifications of a matrix-sentence is generated, the "stored" constituent is substituted in, and another optionally derived sentence is thereby also specified."

2. " suppose that our general meta-theory of grammars supplies, in addition to a specification of the form of grammatical rules, also some complex scheme for "traffic laws" within a grammar, and each grammar has then a "control unit" which directs the order of application of rules, permitting both simple recursions, or reapplications of a rule before moving on, as well as loops in which the path of derivation through the rules curves back on itself to pass through the rules which have already been applied in the generation of certain sentence types"

3. "construct all rules in such a manner that no particular order of application need be assumed, as in the usual version of Turing's formalism for computable functions; each specification of an internal computation (application of a rule) is assumed to apply if and when it can, and only one specification can apply at any one place."

4. " one could retain all the advantages of strict ordering of rules in various places in the grammar but still have relatively simple traffic laws merely by having the latter permit derivations to run through all the rules consecutively to produce a set of (1st order) sentences and then back through all or some of the rules again with whichever strings are still capable of further transformation to less central derived (2nd order) sentence types. This is then a kind of special case of our second suggestion, in which the flow-chart given by the traffic rules is particularly simple: start at #S#, apply consecutively all rules which apply obligatorily and any number of optional rules, at a certain point (or points) return to the beginning and repeat, and give as output any string of terminal symbols which occurs."

Make of all this what you will.

 Aspects introduces recursion in the base, described by Chomsky as allowing the base rules to apply cyclically. p.134

We now allow the rules of the base to apply cyclically, preserving their linear order. Thus, for example, after having generated (1), with #S# in place of S', they reapply to the new occurrence of #S# in the terminal line of the derivation represented by (1). From this occurrence of #S# the rules of the base can generate the derivation represented by (2), with #S# in place of the occurrence of S' in (2). From the latter occurrence of #S#, the same base rules can reapply to form the derivation represented by (3). In this way, the base rules will generate the *generalized Phrase-marker* formed from (1), (2), (3) by replacing S' in (1) by (2) and replacing S' in (2) by (3).

We have thus revised the theory of the base by allowing #S# to appear on the right in certain branching rules, where previously the dummy symbol S' had appeared, and by allowing the rules to reapply (preserving their order) to these newly introduced occurrences of #S#. A generalized Phrase-marker formed in this way contains all of the base Phrase-markers that constitute the basis of a sentence, but it contains more information than a basis in the old sense since it also indicates explicitly how these base Phrase-markers are embedded in one another. That is, the generalized Phrase-marker contains all of the information contained in the basis, as well as the information provided by the generalized embedding transformations.[8]

As for the transformational cycle:

In addition to the rules of the base, so modified, the grammar contains a linear sequence of singular transformations. These apply to generalized Phrase-markers cyclically, in the following manner. First, the sequence of transformational rules applies to the most deeply embedded base Phrase-marker. (For example, it applies to (3), in the generalized Phrase-marker formed by embedding (3) in (2) and the result in (1), as described earlier.) Having applied to all such base Phrase-markers, the sequence of rules reapplies to a configuration dominated by S in which these base Phrase-markers are embedded (to (2), in the same example), and so on, until finally the sequence of rules applies to the configuration dominated by the initial symbol S of the entire generalized Phrase-marker (to (1), in our example). Notice that in the case of (1)–(3), the effect of this convention is precisely what is described in the Transformation-marker (5). That is, singular transformations are applied to constituent sentences before they are embedded, and to matrix sentences after embedding has taken place. The embedding itself is now provided by the branching rules of the base rather than by generalized transformations. We have, in effect, converted the specific properties of the Transformation-marker (5) into general properties of any possible transformational derivation.

The grammar now consists of a base and a linear sequence of singular transformations. These apply in the manner just described. The ordering possibilities that are permitted by the theory of Transformation-markers but apparently never put to use are now excluded in principle. The notion of Transformation-marker disappears, as does the notion of generalized transformation. The base rules form generalized Phrase-markers that contain just the information contained in the basis and the generalized transformations of the earlier version. But observe that in accordance with the Katz-Postal principle discussed earlier (p. 132), it is precisely this information that should be relevant to semantic interpretation. Consequently, we may take a generalized Phrase-marker, in the sense just defined, to be the deep structure generated by the syntactic component.

It is interesting to note that the Aspects syntactic empirical arguments for recursion in the base and cyclic T's were 'negative': There are derivations we don't want to allow. This differs from the most common arguments in the mid tp late 1960s, which were positive: We need the cycle to allow A_B_A derivations.

In Topics (1966) Chomsky again discusses the defects in the LSLT model (observations correctly or incorrectly attributed to Fillmore). He observes that the T-markers he has given have just the desired properties ("the properties Fillmore outlined"):

"That is, singulary transformations are applied to a matrix sentence only after embedding and the only ordering is among singularies." p.62

He then continues:

But the earlier theory of T-markers left open the possibility for ordering of a much more complex sort. It is therefore quite natural to generalize from these empirical observations, and to propose as a general condition on T-markers that they must always meet Fillmore's conditions and have the form illustrated in (19).

As just formulated, this principle appears to be quite ad hoc, but there is another way of saying exactly the same thing that makes it seem entirely natural. Notice that if no singulary transformations apply to a matrix phrase-marker before embedding, and if, furthermore, all embedding involves the insertion of a constituent phrase-marker in a position marked by a dummy element in the matrix structure, then we can, in fact, dispense with generalized transformations entirely. Instead of introducing constituent phrase-markers by embedding transformations, we can permit the re-writing rules of the base to introduce the initial category symbol S, i.e. we can permit rewriting rules of the form A → ...S...

Wherever such a symbol is introduced, we can allow it to head a new base derivation. In short, we can apply the linearly ordered system of base rewriting rules in a cyclic fashion, returning to the beginning of the sequence each time we come upon a new occurrence of S introduced by a rewriting rule. Proceeding in this way, we construct what we can call a *generalized phrase-marker*.

We now apply the linear sequence of singulary transformations in the following manner. First, apply the sequence to the most deeply embedded structure dominated by S in the generalized phrase-marker. Having completed the application of the rules to each such structure, reapply the sequence to the 'next-higher' structure dom-inated by S in the generalized phrase-marker. Continue in this way, until, finally, the sequence of transformations is applied to the structure dominated by the occurrence of S which initiated the first application of base rules, i.e., to the generalized phrase-marker as a whole. Notice that with this formulation, we have, in effect, established the particular formal properties of the T-marker (19) as general properties of any transformational derivation.

Thus, the revised theory is claimed by Chomsky to eliminate the need for an ad hoc stipulation <and maybe another, given my argument about LSLT banning recursion in the base>. Further,

the theory is simpler in eliminating GTs and T-markers.

Interestingly, the first arguments Chomsky gives (In both Aspects and Topics) are based on interface considerations rather than pure syntax. The issue is how syntax connects to semantics. The interface in the LSLT model is the T-marker. There is no DS, in the later sense, since there is no single phrase marker that initiates the transformational derivation. The T-marker indicates how the separate trees in the forest are put together into one tree. Further, there are optional singulary transformations that have an effect on meaning (like negative insertion or SAI). But then we have an argument reminiscent of the syntactic argument: In principle, any points in the derivation could be relevant to semantic interpretation. Could SS be the interface? No. Grammatical relations ((logical) subject of, object of, etc., which Chomsky took to be the core of semantic interpretation, are not available there, because of the effects of transformations such as passive.
BUT, if we eliminate GTs and optional singulary Ts, then the an initial phrase marker exists and is a (pretty) good interface representation. Aspects p.132:

> First, it has been shown that many of the optional singulary transformations of Chomsky (1955, 1957, 1962) must be reformulated as obligatory transformations, whose applicability to a string is determined by presence or absence of a certain marker in the string. This was pointed out by Lees (1960a) for the negation transformation, and by Klima (personal communication) for the question transformation, at about the same time. In fact, it is also true for the passive transformation, as noted in § 2.3.4 of Chapter 2. Katz and Postal (1964) have extended these observations and formulated them in terms of a general principle, namely that *the only contribution of transformations to semantic interpretation is that they interrelate Phrase-markers* (i.e., combine semantic interpretations of already interpreted Phrase-markers in a fixed way).[3] It follows, then, that transformations cannot introduce meaning-bearing elements (nor can they delete lexical items unrecoverably, by the condition mentioned in note 1). Generalizing these remarks to embedding transformations, they conclude also that a sentence transform embedded in a matrix sentence Σ must replace a dummy symbol of Σ. (In the foregoing discussion, adopting this suggestion, we have used S′ as the dummy symbol — this assumption is also implicit in Fillmore, 1963.)

Katz and Postal point out that the principle just stated greatly simplifies the theory of the semantic component, since semantic interpretation will now be independent of all aspects of the Transformation-marker except insofar as this indicates how base structures are interrelated. They have also succeeded in showing that in a large variety of cases, where this general principle has not been met in syntactic description, the description was in fact incorrect on internal syntactic grounds. The principle, then, seems very plausible.

The modified theory (the 'standard theory') instantiates these generalizations p. 135:

Thus the syntactic component consists of a base that generates deep structures and a transformational part that maps them into surface structures. The deep structure of a sentence is submitted to the semantic component for semantic interpretation, and its surface structure enters the phonological component and undergoes phonetic interpretation. The final effect of a grammar, then, is to relate a semantic interpretation to a phonetic representation — that is, to state how a sentence is interpreted. This relation is mediated by the syntactic component of the grammar, which constitutes its sole "creative" part.

The branching rules of the base (that is, its categorial component) define grammatical functions and grammatical relations and determine an abstract underlying order (cf. § 4.4, Chapter 2); the lexicon characterizes the individual properties of particular lexical items that are inserted in specified positions in base Phrase-markers. Thus when we define "deep structures" as "structures generated by the base component," we are, in effect, assuming that the semantic interpretation of a sentence depends only on its lexical items and the grammatical functions and relations represented in the underlying structures in which they appear.[9] This is the basic idea that has motivated the theory of transformational grammar since its inception (cf. note 33, Chapter 2). Its first relatively clear formulation is in Katz and Fodor (1963), and an improved version is given in Katz and Postal (1964), in terms of the modification of syntactic theory proposed there and briefly discussed earlier. The formulation just sug-

Chomsky does acknowledge some potential difficulties. In fn. 32, pp. 221-222, he observes that Topic-Comment structure seems to depend on SS (attributing the observation to Paul Kiparsky. And in fn 9 on p.224, he observes semantic effects of Ts on sentences with quantifiers (an issue that became an intense topic of investigation for at least the next decade and a half). Discussing the claim that DS determines meaning, he says:

> 9. As it stands, this claim seems to me somewhat too strong, though it is true in one important sense of semantic interpretation. For example, it seems clear that the order of "quantifiers" in surface structures sometimes plays a role in semantic interpretation. Thus for many speakers — in particular, for me — the sentences "everyone in the room knows at least two languages" and "at least two languages are known by everyone in the room" are not synony-

The next relevant discussion of the cycle is in Conditions on Ts, which we will discuss much more later:
-The relevant 'cyclic nodes' - domains of application of cyclic Ts - are S and NP in Remarks on Nominalization and S' and NP in Conditions on Ts. <Yeah, I know, he keeps saying S, but he meant NP because he had this weird PS rule S $\rightarrow$ Comp S'>
Conditions on Ts p.241
> "The Insertion Prohibition <<from Aspects; more on this in a minute>>, now sharpened as a special case of the TSC and SSC, is a step toward a stricter interpretation of the cycle: it asserts that once a stage of the cycle has been passed, we cannot introduce material into it from outside under the stated conditions."
Then he proposes further 'sharpening' the principle of the cycle:

> (51) 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
>
> In other words, rules cannot in effect return to earlier stages of the cycle after the derivation has moved to larger, more inclusive domains. We will refer to (51) as the "Strict Cycle Condition."