Crash-Proofness of the Lexicon and Bare Phrase Structure

Miyoko Yasui

Dokkyo University

myasui@dokkyo.ac.jp

1. Crash-Proof Selection Versus Initial Lexical Choice

Frampton and Gutmann (2000, henceforth F&G) propose a very interesting theory of syntax that is claimed to be 'crash-proof': an optimal derivational system from a computational point of view is the one that does not generate ill-formed objects at any stage of derivation. If the derivation starts with some lexical items that are chosen arbitrarily from the lexicon as in Chomsky (1995,1999,2000,2001), it is quite likely to crash. F&G's theory thus dispenses with initial lexical choices and adopts a few syntactic constraints to make lexical items introduced one by one from the lexicon in the 'correct' order. I will discuss that the crash-proofness F&G intend to achieve cannot be efficiently guaranteed within narrow syntax but should be deduced from a native speaker's vast knowledge of lexical items. Instead of pursuing F&G's crash-proof selection or Chomsky's approach with initial lexical choices, I will attempt to replace the standard syntactic tree with a strictly minimalistic graphical representation that can straightforwardly inherit the crash-proofness from the lexicon.

2. Crash-Proof Selection

One of F&G's main points is the 'correct' order of selection. The system guarantees the correct order crucially depending on the following assumptions:

(1) a. Only a lexical item which selects nothing can start the derivation.

b. Satisfaction of selectional requirement is given priority over that of agreement in introducing lexical items. Specifically, given three lexical items X, Y and Z, where X has already been introduced into the derivation, while Y and Z are not, Y is to be introduced into the derivation before Z if a selectional relation holds between X and Y, and an agreement relation holds between X and Z.

Consider the derivation of (2a), which consists of the four lexical items in (2b):[1]

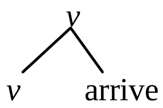
(2) a. Men arrived.

b. {men, [past], *v*, arrive}

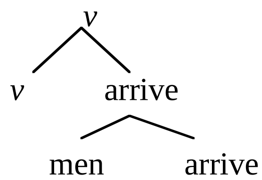
Arrive selects nominal, [past] selects *v*, *v* selects verbal, and *men* selects nothing. Given (1a), *men* is the first lexical item to be introduced into the derivation. *Men* can be selected by the verb *arrive* and it can agree with (or valuate the unvalued f-features of) the tense [past]. Given (1b), *arrive* is to be introduced next. *Arrive* can be selected by *v*, and this is the only option at this stage. The tense [past] can select *v* and agree with *men*. Given (1b), [past] is introduced as the head selecting *men*. Finally, the agreement requirement of [past] is satisfied by the internal merger of *men* into its specifier position.

Note that (1a) is necessary in blocking the derivation from starting with [past] or *v* rather than *men*. If *v* is introduced into the derivation first, it merges with *arrive* next, forming structure (3a):

(3) a.



b.



Then, cyclicity precludes *men* from being introduced into the derivation as in (3b).

One obvious problem with this account is that (2b) does not exist in F&G's system; lexical items are to be freely chosen from the whole lexicon at each stage of derivation.

[1] For expository convenience, I assume with Chomsky (1999) that there are two light verbs v and v* selecting intransitive and transitive verbs, respectively, though F&G analyze (2a) without v. The difference does not affect the discussion here.

Lexical items that satisfy (1a) are mostly nouns.[2] Thus, the first step is not selecting *men* from the set in (2b) but selecting some noun non-deterministically from the whole set of nouns. Similarly, the next step is not selecting *arrive* from (2b) but selecting some verb non-deterministically from the lexicon. The verb in question need not be intransitive; merging *men* with a transitive verb such as *hit* can lead to a well-formed structure if one more appropriate noun is introduced into the derivation. It is true that (2a) is derivable in one 'correct' order but only on the basis of accidental and time-consuming selections of the four lexical items from the whole lexicon.

The situation seems no better than the massive overgeneration of initial lexical arrays in Chomsky's system, most of which are doomed to fail, as pointed out by F&G. In F&G's system, several (and often a great number of) lexical items are searched as a candidate for each lexical item that will eventually constitute a sentence, but none of them are actually chosen except for one. On the other hand, a possible algorithm instantiating Chomsky's system might be to list up all the possible subsets of the lexicon starting with size *n*, with size *n*+1 and so on, where *n* is the length of a shortest possible sentence. Derivations from most of them crash, but adopting some appropriate version of (1a,b) and other auxiliary assumptions could help restrict overgeneration.

Next, consider the derivation of the transitive sentence in (4a) with its components in (4b):

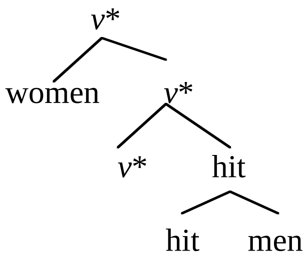
(4) a. Women met men.

b. {[past], v*, hit, women, men}

(4a) has structure (5) at an earlier stage of its derivation:

[2] Besides nouns, the verb rain, selecting nothing, presumably satisfies (1a).

(5)

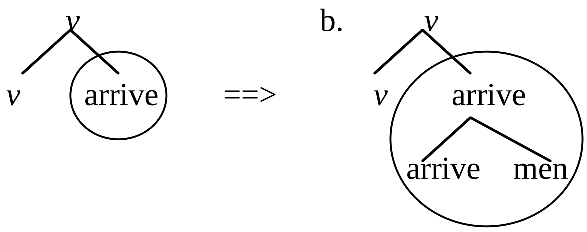


In (5), *hit* selects *men*, and if the relation between the light verb *v**and its external argument is selection, *v** selects *women*. Both *men* and *women* are nouns and thus can start the derivation. It does not matter which noun does. If *hit* is introduced after one of the nouns, say *men*, (5) can be derived without a violation of cyclicity. The problem is that *v** can be introduced as the selector of *men* before *hit*. If *men* is selected by *v**, introducing *hit* in the next step would violate cyclicity. One might try to circumvent this problem by claiming that *v** only selects verbal and its relation with an external argument is something other than selection. This amounts to adding further complication to (1b). (1b) or its further extension is almost a restatement of what the correct order of merger is, and if Chomsky's system with initial lexical arrays is supplemented with similar assumptions, it will become as crash-resistant as F&G's.

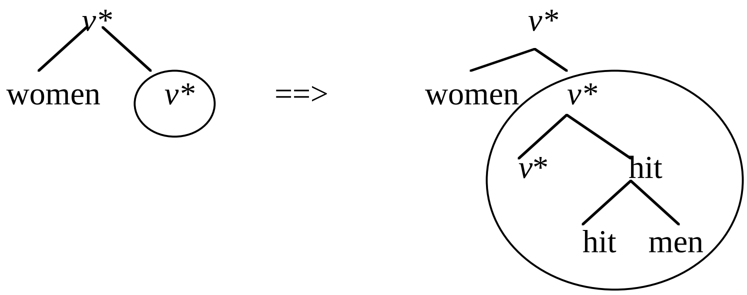
3. Violations of Cyclicity and Graphical Representations of Syntactic Structure

We have seen that F&G's system needs (1a) and some appropriate extension of (1b) to make syntactic derivations proceed in an unambiguous manner without violating cyclicity. The unwanted violations of cyclicity are illustrated again below:

(6) a.



(7) a.

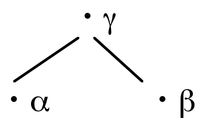


Once (6a) and (7a) are formed, cyclicity disallows the circled part to be expanded as in (6b) and (7b), respectively.

Let's find out which part of the circled structures is problematic. Introducing a new lexical item into the structure already formed should be allowed since it is indispensable part of structure-building. Thus, *men* in (6b) and the lower *hit* and *men* in (7b) are not problematic at all. Moreover, the lower *arrive* in (6b) and lower *v** have been present as the lexical heads in (6a) and (7a), respectively. Offending are the upper *arrive* in (6b), and the upper *v** and *hit* in (7b), all of which are non-terminal symbols. If these offending non-terminal nodes are somehow eliminable on principled grounds, the violations of cyclicity found in (6) and (7) probably will cease to be an issue; so will the crash-proofness of selection.

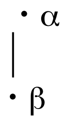
Note, first of all, that non-terminal symbols have been postulated, by definition, in rewriting rules/merging operations and their graphical representations upheld since the beginning of generative grammar: the combination of constituents is represented as a node distinct from them, as in (8):

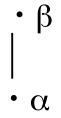
(8) a. • α • β b.



α and β in (8a) are inputs to merger (or the righthand part of a rewriting rule). Merging α and β has been assumed to result in (8b), where the new node γ is introduced along with two directed edges connecting it with α and β. Chomsky (1995) argues that the label of γ can be identified with that of a or that of b in a principled manner; still, γ is a node distinct from α and β.[3]

The above representation of syntactic structure is not mandatory. Alternative and simpler graphical representations are (9a,b) corresponding to the two cases of (8b) with the head being a and b, respectively:

(9) a. 

b. 

Merging α and β is graphically represented just by drawing a directed edge from one of them toward the other. Let us examine (8b) and (9a,b) from a graph-theoretic point of view. A graph, denoted G=(V, E), consists of a set of nodes V and a set of ordered pairs of nodes E.[4] (8b), (9a,b) are defined as follows:

(10) (8b): G1=(V1, E1) $V1= \{\alpha, \beta, \gamma\}$. $E1 =\{<\gamma, \alpha>, <\gamma, \beta>\}$

the label of γ is equal to that of α or β.

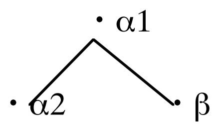
(9a): G2=(V2, E2) $V2=\{\alpha, \beta\}$ $E2=\{<\alpha, \beta>\}$

(9b): G3=(V3, E3) $V3=\{\alpha, \beta\}$ $E3= \{<\beta, \alpha>\}$.

(8b) and (9a,b) differ in their representation of headedness. In (8b), a non-terminal node γ is added to the node set V1, two ordered pairs $<\gamma,\alpha>$ and $<\gamma,\beta>$ constitute the edge set E1, and the label of γ is identified either by that of α or β. In (9a,b), on the other hand, the node sets V2 and V3 are minimal, containing just α and β, and either of the two possible ordered pairs $<\alpha, \beta>$ and $<\beta, \alpha>$ constitutes the edge set. Note that $<\alpha, \beta>$ is generally defined as

---

[3] Chomsky (1995: 245) recognizes the necessity to distinguish a label for a lexical item and the same label for the structure headed by the lexical item in terms of subsctipts as follows:

.(i) 

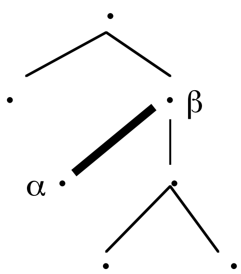
α1 and α2 in (i) are distinct nodes with distinct labels.

[4] Instead of using arrows to express ordered pairs, downward lines have been used in standard tree diagrams such as (8b), and this tradition is followed here.

$\{\{\alpha\}, \{\alpha, \beta\}\}$, and it looks rather similar to Chomsky's (1995:244-245) definition of the object formed from a and b of the type α: $\{\alpha, \{\alpha, b\}\}$. For expository convenience, Chomsky continues to employ the graphical representation of the form (8b) for the object $\{\alpha, \{\alpha, \beta\}\}$, acknowledging that (8b) is more complex than is absolutely necessary. If $\{\{\alpha\}, \{\alpha, \beta\}\}$ is adopted instead of $\{\alpha, \{\alpha, \beta\}\}$ as the definition of the object formed from α and β , the discrepancy between the formal definition and its graphical representation will disappear, which seems to be a desired result.

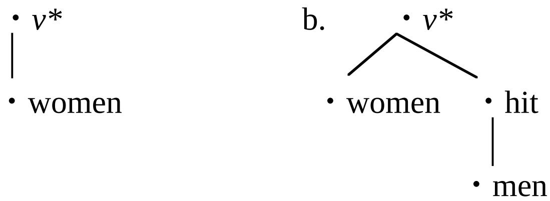
If the strictly minimalistic system of graphical representation is taken along the above line, introducing a new lexical item will not distort the structure already formed, and hence the order of lexical introduction will become a trivial issue. Graphically, it just adds a new edge (the bold line in (11)) connecting the new lexical item α and one of the nodes in the graph already formed (β) : the graph consisting of the solid lines is retained as a subgraph of the graph formed by the introduction of α:

(11)



Set-theoretically, α is added to the node set of the graph already formed, and $\langle\alpha, \beta\rangle$ (or $\langle\beta, \alpha\rangle$) is added to the edge set. Going back to the case of cyclic violation described in (6), the strictly minimalistic system depicts (6a,b) as (12a,b), respectively:

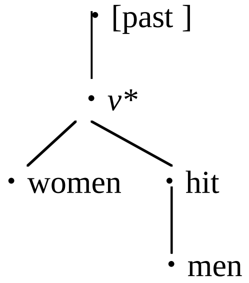
(12) a. b.



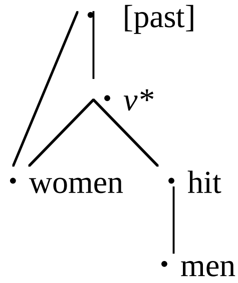
(12b) retains (12a) as its subgraph. It is easy to verify that introducing the four lexical items in any order will inherit the graph already form as its subgraph. In this way, the alternative system succeeds in eliminating (1a).

To see if the priority of selection over agreement stated in (1b) can be likewise eliminated, consider the derivation from (12b) to (13a), where [past] is introduced as the selector of *v**:

(13) a.



b.



Women should agree with [past]. Suppose that this relation is expressed by the downward edge from [past] to *women* as in (16b). (16b) inherits (16a) as its subgraph, and(16b) will be produced if its derivation starts with [past] and *women* rather than *men*, which might suggest that (1b) be given up. Nevertheless, I will retain (1b) in a different form , arguing in Section 5 for the conception of initial syntactic structures as subgraphs of the lexicon, which is claimed to be a huge connected graph expressing the selectional properties of lexical items.

4. Preference of Merge over Move and the Full Use of Lexical Items

Chomsky (MI) argues for the phase-based initial lexical choice based on (14) and (15):

(14) a. There is likely to be a proof discovered.

b. *there is likely a proof to be discovered

(15) It was decided [PRO to be executed at dawn]]

(14a) has had (16) at an earlier stage of its derivation:

(16) [to$_{def}$ [be a proof discovered]]

The EPP-feature of the defective infinitive tense requires that something occupy its specifier position. Two options are available: merge *there* or move *a proof*. Here, (17) comes into play, forcing (16) to expand into (18a) rather than (18b):

(17) Simple operations (Merge) preempt more complex ones (Mmove or Merge plus Agree)

(18) a. [there to_{def} [be a proof discovered]]

b. [a proof to_{def} [be (a proof) discovered]]

Hence, (14a) is derivable from (18a), but (14b) is not. (17), however, incorrectly blocks the movement of PRO to the specifier of the control t*o* in (15) since the expletive *it* is available; *it* should merge with *to be executed RPO at dawn*. To overcome this difficulty, Chomsky introduces the idea of phase-based lexical choice: from the set of lexical items initially chosen to build a whole sentence, some subset is extracted and placed in active memory for building up a phase of the sentence. Given the phase-based lexical choice, *it* is not available in the embedded phase of (15); and it is derivable.

Note that a subset of lexical items placed in active memory needs to be used up before the computation proceeds, as Chomsky (2000) says. This condition, however, appears to weaken the motivation for (17). Going back to (16), t*here* is in active memory and it must be used before the computation goes on to the next phase. Simply for this reason, *a poof* cannot move; if it moved, *there* would go nowhere within the phase. Chomsky (2000:103-104) argues that (17) is crucial in deducing the order of object-subject in so-called Object Shift languages. Moreover, some amount of look-ahead is involved in the account of (14a,b) based on the requirement that lexical items be used up in each phase. Still, this requirement reduces the necessity of (17).

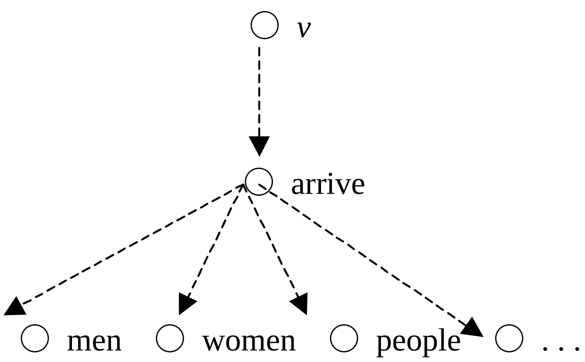
5. Lexicon as a Graph and Syntactic Structure as its Subgraph

There have been pointed out some problems in F&G's theory of crash-proof selection as well as in Chomsky's phase-based theory of structure-building. In this section, I will try to accommodate their insights into the graph-theoretic approach suggested in Section 3 and to work out some of the problems. The important point has been that the graphical representation adopted here has only lexical items as its nodes. Non-terminal symbols such as V' have been eliminated since they are largely responsible for cyclicity violations and the issue of lexical introduction.

The question arises how such a graphical representation is obtained It might be formed by introducing lexical items one by one from the lexicon as F&G claim, or if the lexicon itself is a graph with lexical items as its nodes and their selectional relations as its edges, which is plausible, the graphical representation of syntactic structure can be regarded as some appropriate subgraph of the lexicon.[5] By definition, the lexicon is about lexical items only; there are no entries for non-terminal symbols. Each lexical item is related to some other lexical items; if a lexical item were totally dissociated, it would never be selected in syntactic derivation. The relation at stake might be semantic (s-selection) or categorial (c-selection). It might also include valuation of some uninterpretable feature (agreement), but I will assume it does not here, and I come back to this issue shortly. In this way, it is reasonable to represent each lexical item as a node of the graph L(exicon) and its selectional relations with other lexical items in terms of directed edges.

For instance, *arrive* selects nominal and is selected by the light verb *v*. Then, the graph L involves the following as one of its numerous subgraphs:

(19)

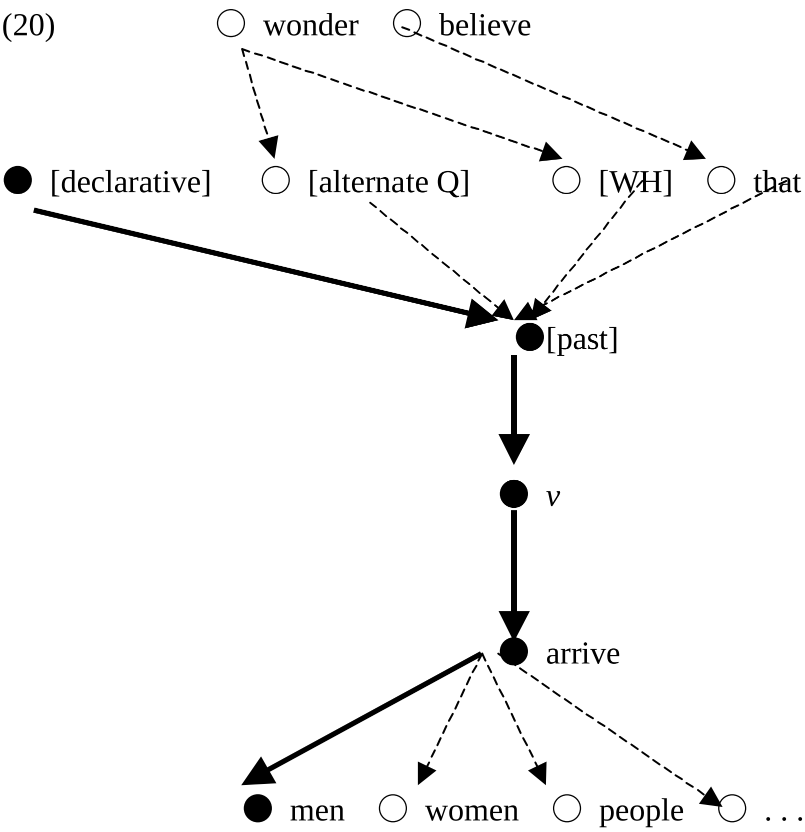


[5] The former view is taken in Yasui (to appear), which is an attempt to explain the correlation between the head-parameter and the overt/covert distinction in wh-movement. The account is compatible with the latter view.

I will not go into the question of whether *arrive* is connected to all nominals or some semantically appropriate nominals, which is a matter of convergence or gibberishness.

Arrive is connected indirectly with other lexical items in L as follows:

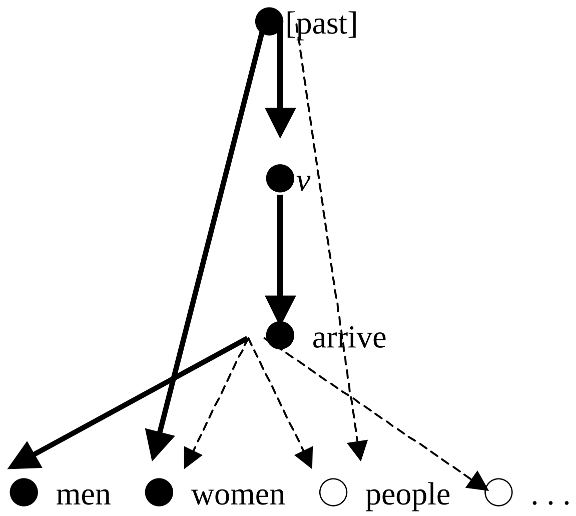
(20)



Note that the subgraph of (20) consisting of the filled circles and bold arrows is the graphical representation of (2a) before the movement of *men*. More generally speaking, any well-formed structure before it is subject to movement operations is some subgraph of L. The converse does not hold, however; not all subgraphs of L will converge in syntax. For one thing, a subgraph needs to be rooted by one of the specified lexical items: [declarative], [alternative Q] and [WH]. [declarative], denoting the COMP of root declarative clause, is a special node in that it has no mother node or incoming edge. [alternate Q] and [WH] can head root or embedded wh-questions, but t*hat* can head only embedded declarative clauses. Another condition on convergence is that the dead end nodes (those that have no outgoing arrows) of the subgraph must also be dead ends in L. If a dead end node in the subgraph has outgoing arrows in L, it means that the lexical item the node expresses should select something but it does not in the subgraph. As has been mentioned in connection to F&G's assumption (1a), nouns do not select obligatorily; nouns are dead ends in L. (1a) remains in the theory here in this way, though the necessity to introduce lexical items in a particular order has been completely eliminated.

Since L is huge, picking up some subgraph from it should be a non-trivial task, but the condition on the root node might reduce the difficulty to some extent. Moreover, it is largely a matter of our thinking rather than syntax. Suppose that a subgraph G=<V, E> is somehow extracted from L. It is ensured that the selectional requirements of all the lexical items in V are satisfied since they have been extracted from the lexicon so as to meet this condition. The remaining job is to satisfy agreement requirements in G if any by Move in syntactic derivation. In (20), no arrow is assumed to exist between [past] and *men*. It is because [past] can agree with a lot of nouns and if its agreement possibilities are expressed in L by arrows as in (21), something like "*women arrived men" results. :

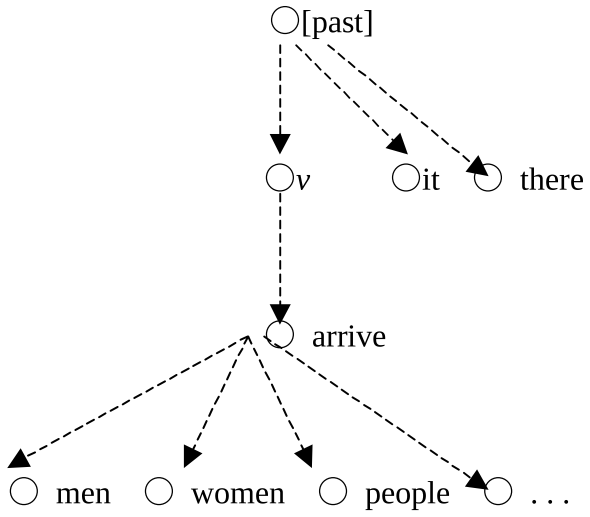
(21)



In (21), *men*, which is the argument of *arrive*, should agree with [past] but it does not. If the directed edges of L express only selectional relations among lexical items, excluding agreement, subgraphs that are doomed such as (21) are never extracted. This architecture of L has essentially the same function as F&G's assumption (1ii): satisfaction of selectional requirement is given priority over that of agreement in introducing lexical items. It also corresponds to preference of Merge over Move in Chomsky's theory. The important difference is that F&G's and Chomsky's assumptions are part of syntax but it is part of the lexicon here. In other words, a native speaker's knowledge of the lexical items in his/her language ensures (1ii) or preference of Merge over Move.

Though nouns are not connected to tenses in L, expletive *there* and *it* should be as follows; otherwise, they would never be introduced into syntactic structure:

(22)



Going back to (14a,b) and (15) repeated below:

(14) a. There is likely to be a proof discovered.

b. *there is likely a proof to be discovered

(15) It was decided [PRO to be executed at dawn]]

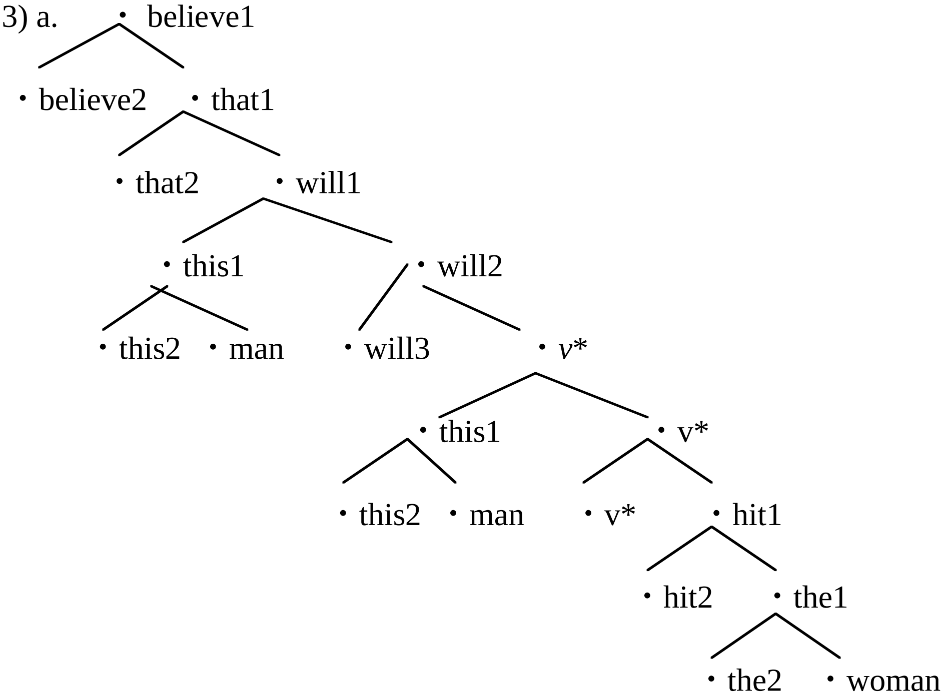
this pattern of data can be explained if extraction of a subgraph is phase-based, as Chomsky assumes.[6] Preference of Merge over Move can be eliminated. Instead, the condition that all the lexical items must be used up before the computation proceeds is subsumed under the conception of structure-building here: a subgraph of L is submitted to the syntactic computation and all the lexical items in it are necessarily used.

[6] A sentence often contains more than one occurrences of some light verb and COMP1. If a subgraph is extracted to build a phase, it contains no more than one light verb or COMP, hence, the phase-based extraction appears to make the system simpler.

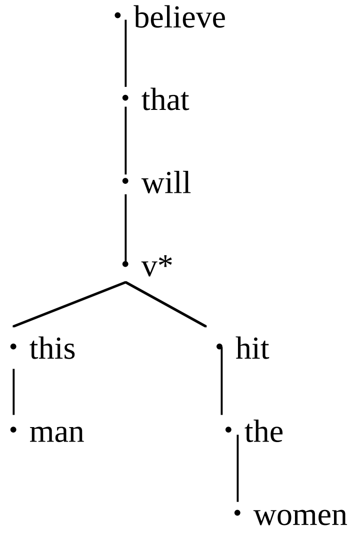
6. Defining Some Fundamental Structural Relations

How are the two important relations Sister and Immediately-Contain in Chomsky (2000) to be defined in the alternative theory here? Consider the standard tree diagram of the phrase "believe that this man will hit the woman" in (23a) and its alternative version in (23b):

(23) a.



b.



The movement of *this man* is expressed as two occurrence of it in (23a) and the additional edge with this in (23b). The movement of object is ignored here for expository convenience.

The set-theoretic definitions of the edge set of (23b) is (24a) and equivalently (24b).

Sisterhoood in the standard sense can be defined by adjancent nodes or ordered pair in the alternative theory. In (16a), *believe2* and the subtree rooted by the node *that*1 are Sisters.[7] In (23b), *believe* and the subgraph rooted by its adjacent node *that* are Sisters. One difference is that the subject *this man* and *will* are Sisters in (23b) according to this definition since *this* and *will* are adjacent nodes, but they are not strictly Sisters in (23a) though they m-command each other. This might or might not be a problem.

The notion of Immediately-Contain can also be defined in terms of adjacent nodes. In (23a), *will*2 Immediately-Contains *will*3and the subtree rooted byhit*t*1 in (23a). In (23b), the graph rooted by *hit* is a proper subgraph of that rooted by its superjacent node *will*. The Immediately-Containing relation between *will*2 and *will*3 in (23a) presumably has no significance; hence, the absence of the corresponding notion in (23b) is not problematic.

The c-command relation holding between *this man* and *the woman* in (23a), which is important at least in interpretation, can be translated into (23b) in terms of the same definition. One type of c-command relation that is present in (23a) but absent in (23b) is that holding between a head and a phrase other than its complement. In particular, *will*3 c-command *the woman* in (23a), but the corresponding relation does not exist in (23b), which contains only one instance of *will*. This is unlikely to be problematic, since the notion of c-command is significant only for phrases such as a reflexive and its antecedent.

Note that the definition of closest heads such as *will* and *hit* in the phrase under consideration requires some complication in (23a), but they are simply defined as two adjacent nodes in (23b). Actually, an ordered pair (or two adjacent nodes) constitutes the fundamental structure in the alternative system of graphic representation here. This appears to be an

[7] The asymmetric c-command relation holding between nodes rather than constituents plays a crucial role in Kayne's Liniear Correspondence Axiom. It is not clear if relations holding between nodes rather than constituents are syntactically important.

advantage over the standard tree representation since the notion of closest heads is important in head-movement and searching for relevant f-features by the probe among others.

7. Concluding Remakrs

References

Balakrishnan, R. and K. Ranganathan (2000) *A Textbook of Graph Theory.* Springer.

Chomsky, N. (1995) *The Minimalist Program*. The MIT Press.

_____ (1999) *Derivation by Phase: MIT Occasional Papers in Linguistics no.18*.

_____ (2000) "Minimalist Inquiries: The Framework." In *Step by Step: Essays on Minimalist Syntax in Honor of Howard Lasnik*, eds. Roger Martin, David Michaels, and Juan Uriagereka. MIT Press, Cambridge.

_____ (2001) "Beyond Explanatory Adequacy." ms., MIT.

Frampton, John and Sam Gutmann (1999). "Cyclic Computation, a Computationally Efficient Minimalist Syntax." *Syntax* 2, 1-27.

_____ (2000) "Crash-Proof Syntax. " ms., Northeastern University.

Hopcroft, J. E. and J. D. Ullman (1979) *Introduction to Automata Theory, Languages, and Computation*. Addison-Wesley.

Kayne, R. S. (1994) *The Antisymmetry of Syntax*. The MIT Press.

Yasui, M. (2002) "“A Graph-Theoretic Reanalysis of Bare Phrase Structure Theory and its Implications on Parametric Variation.” ms., Dokkyo University. (http://www2.dokkyo.ac.jp/~esemi003/publications/graph_v1.PDF)

_____ (to appear) “A Graph-Theoretic Reanalysis of Bare Phrase Structure and Wh-movement." *Proceeding of LP2002*. Charles University Press.