

Chomsky: On Phases

Joost Kremers
University of Frankfurt, Germany
j.kremers@em.uni-frankfurt.de

May 12, 2005

This text originated as a handout for a class at the German Language and Literature Department of the University of Frankfurt. Note that although the text is mine, I have benefited much from the discussions with and ideas and solutions offered by the participants in this class, most notably Eric Fuß, Patrick Brandt, Marina Stoyanova, Andreas Runkel, and by some discussion outside this class with Iris Mulders. The version of the “On Phases” paper that I used as a basis is a version of March, 2005.

1 Overview

To summarize the most important changes to the model introduced in the paper:

- There is now only one computational cycle. This cycle builds structures until a phase has been created.
- When a phase has been created, transfer takes place: the phase is sent to the C-I and SM (sensorimotor) systems, where it is interpreted.
- As a result, the level of representation LF no longer exists. In effect, this means that there can no longer be syntactic operations that take place “at LF”.
- T does not have Agree features in and of itself, which means it can no longer act as a probe. Instead, T inherits its Agree features from C. As a result, it is C that ultimately initiates the Agree relation that values the subject’s Case feature and optionally raises the subject. Subject raising goes to SPEC-T, not SPEC-C, however, because C’s Agree features are passed on to T.
- In general, Merge is made possible by an *edge feature* (EF). On the phase head, EF can effectuate internal Merge (IM) of any element in the clause (as long as that IM is not excluded for other reasons). As such, EF is responsible for A’-movement to SPEC-C.
- The relation v^*-V is equivalent to the relation C-T. Therefore, v^* also has EF and Agree features, and as in the case of C-T, the Agree features are inherited by the phase head’s complement, in this case V. As a result, it is assumed that object agreement does not take place in SPEC- v^* , but in SPEC-V.
- In short: all operations are initiated by the phase heads (PH). A PH has two types of features: a) Agree features, which are inherited by the head selected by PH, and which mediate agreement; and b) an edge feature, which is responsible for raising to SPEC-PH.

- The A-A'-distinction is now defined derivationally: a position created after Merge initiated by EF is an A'-position, typically SPEC-C and outer SPEC-v*. Other positions are A-positions.
- A-movement is initiated by the phase heads, but “by proxy”: it takes place only when T has inherited the Agree features from C, and V from v*. Note that this means that T can only attract something to its SPEC after C has been merged: it needs to inherit C’s Agree features before it can do this.
- A'-movement is initiated by the phase head directly, and always through EF. This means that it is no longer assumed that the C-domain contains uninterpretable *wh*-, Top-, Foc-, Q-features, etc. EF can in principle attract any element, and the position in which an element ends up determines its final interpretation.
- This means that the computational system *overgenerates*: it can create structures that do not receive any interpretation at the interfaces. In fact, convergence no longer seems an absolute criterion. The computational system is rather free in its ability to generate structures, and “deviance” is determined at the interfaces.
- Once an element has A-moved to the position where its features are valued, it becomes invisible to EF. As a result, there is no A'-movement from SPEC-T. A'-movement therefore originates from a lower position. E.g. in subject-*wh*, C attracts *wh* from SPEC-v*, not from SPEC-T. In fact, C and T *both* attract *wh* from SPEC-v*, and do so *in parallel*. With successive-cyclic movement, C can also attract *wh* out of an intermediate position, e.g. after one or two steps in the A-movement, but before or simultaneously with further A-movement steps.

2 The Summary

Text in times constitute the summary, with much text quoted directly from Chomsky’s paper. Text in sans serif font are explanations, interpretations and comments from me.

- The Faculty of Language is “an organ of the body”. Obviously, FL is not an organ in the normal sense of the word. Rather, it is a cognitive system, such as vision, where it should be noted that it would appear that language is more spread throughout the brain than vision, and is in a way a “higher level” cognitive system.
- Three factors enter into the growth and development of FL: I. external data; II. genetic endowment; III. principles of structural architecture and developmental constraints not specific to FL.
- It seems unlikely that FL can be attributed entirely to months or years of experience, rather than to millions of years of evolution or to principles of neural organisation.
- Two levels of explanation: explanatory adequacy is reached when we have a theory of the genetic endowment that accounts for FL. An explanation is principled when we can reduce it to factor III and to the so-called interface conditions.
- One aspect of a principled explanation that we would expect for a computational system such as language consists of concepts of computational efficiency. In general, it is expected that conditions holding for FL are of a more general nature, and can be found to work in other domains (cognitive systems) and other organisms.

- Strong Minimalist Thesis (SMT): language is an optimal solution to interface conditions that FL must satisfy. Any departure from SMT is a possible indication that the theory contains some superfluous descriptive mechanisms.
- One question is whether the two interfaces contribute equally to the design of FL. There is reason to believe that C-I is more principled (language is poorly designed for communication). There may be devices at the SM interface that are used to overcome some of the communicative problems.
- Conditions imposed by the C-I interface enter into principled explanation in a crucial way, while mapping to the SM interface is an ancillary process.
- Two linguistic levels are assumed to be indispensable: the interface levels that are accessible to SM and C-I. The EST/Y-model postulated three further levels: d-structure, s-structure and LF, each requiring an additional cycle of operations. According to SMT, these levels should not be needed, and preferably even unformulable. I have always thought that LF was the interface level to C-I. What Chomsky probably means is that LF has become a locus for all sorts of syntactic operations, e.g., covert movement, reconstruction, and that this violates SMT.
- The most elementary property of language is that it is a system of discrete infinity consisting of hierarchically organized objects. Chomsky calls this property “an unusual one in the biological world”. I believe that it is only the infinite nature that is really unusual: the cortex is essentially a hierarchical system, e.g., a cognitive system such as vision is best described in the form of a hierarchy. As far as I know, however, the visual system is not infinite; instead, the hierarchy is fixed, so to speak. Infinity in language is created by *recursion*, i.e., the possibility of including an XP inside an XP of the same type, and by what might be called *creativity*, the fact that the hierarchy is *not* fixed but flexible, at least at certain points. Interestingly, both these properties seem to only exist at the phase level. Inside a phase, the structure (hierarchy) *is* fixed. (I do not believe that the possibility to “skip” categories, e.g. in the C-domain, or adjunction, which extends categories, essentially changes this.) If this is correct, then it seems that the essential property of language lies not within narrow syntax, but at a higher level of cognition.
- Such a system requires an operation that takes n syntactic objects (SOs) and constructs a new SO from them. Call this *Merge*. (A production model would seem to benefit more from an operation such as *Expand*, which takes a higher-level SO and expands it into (possible) lower level SOs. Rewrite rules, anyone?) Since Merge is unavoidable in a system of hierarchic discrete infinity, we can say it “comes free”.
- The emergence of unbounded Merge in human evolution, probably a sudden event, would instantaneously have yielded a wide range of new abilities, which might account for the dramatic and sudden changes found in the fossil record.
- Restriction of computational resources probably limits n for Merge to two, i.e., Merge creates binary structures only. Other conditions support this conclusion: minimal search, perhaps interface conditions (requirements of linearisation); predicate-argument structure.
- Efficient computation seems to require a no-tampering condition (NTC): Merge of X and Y leaves the two SOs unchanged. Merge of X and Y can be taken to create the set $\{X,Y\}$. This means that Merge is invariably to the edge, and that we can postulate the *inclusiveness principle*, i.e., there are no bar-levels, traces, indices, etc.

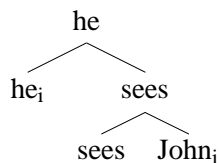
- Alternatively, Merge could create the ordered pair $\langle X, Y \rangle$. The issue is whether linear order plays a role at the C-I interface, or whether it is restricted to SM. The general assumption is that the latter is the case. Hence the assumption that Merge creates a set, not a pair. Note that Kayne's LCA *does* apply to narrow syntax, hence in an antisymmetric model, linear order *is* defined there.
- (Unary) Merge together with a language with the simplest possible lexicon yields arithmetic in the familiar way.
- For a lexical item (LI) to enter into a computation, merging with some SO, it must have some feature allowing this operation. Call this the *edge feature* (EF). (Note that later in the paper, Chomsky defines the edge feature more specifically.) An LI without EF is an interjection. When an LI is merged with an SO, yielding $\{LI, SO\}$, SO is its complement. The fact that Merge iterates without limit is a property at least of LIs — and optimally, only of LIs. EF articulates the fact that Merge is unbounded, that language is a recursive infinite system. The remark “a property *at least* of LIs — and optimally, *only* of LIs” probably refers to the idea that Merge is unique to human language, and is not found in other cognitive or biological systems.
- Reliance on iterable Merge as the sole computational operation of narrow syntax eliminates, as unformulable, the notions d-structure and s-structure, and three of the compositional cycles of EST. I assume Chomsky means that with iterable Merge, levels such as d-structure and s-structure would have to be stipulated: one would have to choose an arbitrary point in the derivation and name it d/s-structure. Such a stipulation would violate SMT.
- When Y is merged to X (this formulation implies that Merge is asymmetrical; it is, of course, given that one of the two elements projects), either Y is not part of X (*external Merge* EM), or Y is part of X (*internal Merge* IM). IM yields two copies of Y in $\{X, Y\}$, one external and one internal to X. IM is *Move* under the copy theory of movement. The confusion about the copy theory of movement Chomsky refers to seems to stem from his terminology. To me, the word *copy* implies a new object, identical to but distinct from the original object. For Chomsky, however, *copy* does not seem to mean that. Instead, a copy seems to mean the *same* object, remerged at another location, creating a configuration in which *one* object occupies two locations in the tree.

Note that Chomsky does not mention the Numeration anymore. So we no longer have a system in which all lexical items are first selected from the lexicon. Apparently, at each instance of Merge, a new item is selected, either from the lexicon or from the SO to be merged.

- If the means of language are fully exploited by the interface systems, then we would expect the two types of Merge to have different effects at the interfaces. At the phonetic interface, they obviously do: IM yields the ubiquitous displacement phenomenon. At the semantic interface, EM yields generalized argument structure, IM yields discourse-related properties. The formulation here almost seems to suggest that the correlation between the duality in Merge and the duality in the C-I system is purely accidental. But wouldn't it be more likely that the two developed in parallel, or one as a result of the other? Note also that the discourse-related properties that IM yields are *communicative* properties. Earlier, Chomsky argued that mapping to the SM interface was “ancillary”, and that language was not optimized for communicative purposes. That would suggest that IM may not have been put to use at all in the earliest forms of language, and only became operative when language started to be used for communication. As such, IM does seem to have some sort of secondary status, which isn't too surprising of course, given that IM can only take place *after* EM has taken place. (It might be argued that the effects that IM yields are simply part of some *language of thought*, just as generalized argument structure. But I

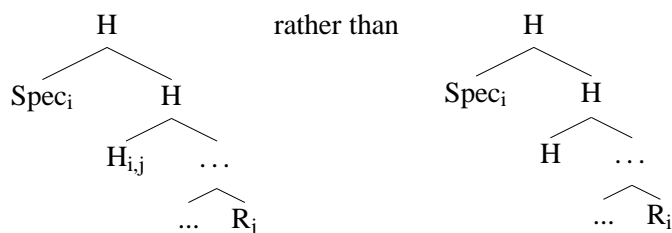
believe they really *are* communicative: topic, focus, *wh*, etc. all seem to require what is sometimes called a *theory of mind*, the notion that others have minds and that their state of knowledge and focus of attention is different from one's own. In a language of thought, new information (focus) or topic of conversation seem meaningless. *wh* may theoretically be of use, but *wh in situ* would seem to suffice: IM of *wh* seems more a matter of marking *wh* for the listener rather than expressing *wh* itself.)

- It is generally assumed that EM comes free. The model as described here shows that contrary to past assumptions, IM comes free as well. Note that Chomsky is not talking about some form of economy here. Economy in the theory has been long abandoned. Chomsky is talking more about a sort of meta-theory, a theory of theory-formation. If the theory takes Merge to be the basic structure-forming operation, then both EM and IM are inevitable: they need not be stipulated, and they cannot be eliminated from the theory without stipulation.
- Each SO enters into further computations. The information about the SO that is relevant for these computations is presumably encoded in a single designated element, the *label*. The label selects and is selected in EM, and functions as the *probe* in operations internal to SO, Agree or IM.
- A natural interpretation of the notion “edge” can capture some of the properties of “tucking in”, taking the “edge” to be the position as close as possible to the probe. Note that this seems a bit problematic in light of the NTC: tucking in requires alteration of the SO, no matter how one formulates it.
- We thus have two syntactic relations: (A) set-membership and (B) probe-goal relations. (A) yields *term-of* and *dominate*. Adding *sister-of* (As a derived notion, I guess) we obtain *c-command* and *identity* (as in *sister-of(sister-of(X)) = X*. This of course requires binary branching trees.) It is not clear whether *c-command* plays a role within the computation to the C-I interface, but let us assume it does not. Note that Chomsky does sort-of assume the LCA (which requires *c-command*) but, unlike Kayne, takes it to be a requirement of the phonological interface, hence not operative in narrow syntax.
- *C-command* is generally assumed to be required in binding theory (BT). If so, BT would be at the “outer edge” of the C-I interface. (Does that mean it is part of C-I, rather than of syntax?) *C-command* may not be required for BT, though. Condition C can possibly be accounted for on the assumption that the *c-commanding* pronoun is a probe. More or less like this:

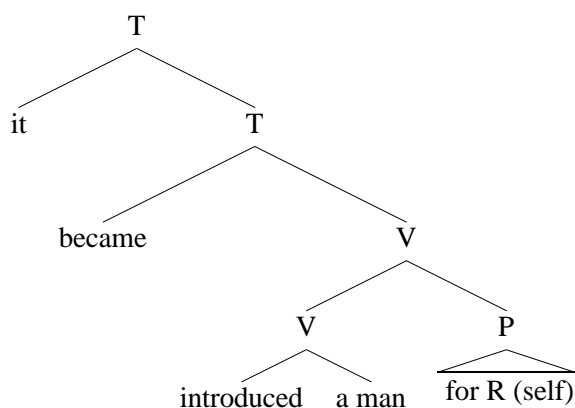


But there are still many questions here: what is *he* a probe for? What would the structure look like exactly? Why could the Agree relation not be established? How does this conclusion relate to the assumption made later that all operations are initiated by phase heads, i.e., that only phase heads can be probes?

- Condition A may be reformulated (for bare subject-oriented reflexives) if we take the head as a probe for the reflexive (with expository indices indicating relevant relations):



The crucial test here is long-distance agreement: a structure of the form [H...XP...R], where H and XP agree, XP does not c-command R, and R is in the minimal search domain of the probe H:



Here, *a man* does not c-command R, but both are goals of the probe that heads the construction (*became*, I assume). In this case, R is a bare form reflexive, meaning it is in an agreement (probe-goal) relation with H, though not c-commanded by its antecedent XP. Trying to make this explicit: R does not agree with *a man*, but it is probed by *became*, requiring an R form? Note that this tree seems to have some problems: it is not clear to me how [introduced a man] can be a constituent here. Isn't the PP an argument of the verb as well? Wouldn't the surface structure then be derived by object shift followed by V movement or something similar? And wouldn't that lead to a configuration in which "a man" does c-command the R element?

- The probe-goal framework is assumed, including Multiple Agree: the probe agrees with goals in its domain as far as a goal with no unvalued features, which blocks further search. Features of the goal (including structural Case) value all matched elements (probe, participles i.e., "intermediate" goals) and the option of raising the goal may or may not be exercised — and if exercised will be all the way to the probe. Multiple Agree in Hiraiwa's framework is instantaneous: an Agree relation is established between the probe and all the goals at once, the lowest one having the possibility to raise.
- Iterated Merge accounts for three of the EST/Y-model's compositional cycles, but two are still unaccounted for: the mappings to the semantic and to the phonological interface (since they are not driven by Merge). These are incorporated (and LF eliminated), if at various stages of computation there are Transfer operations: the SO already constructed is handed to the interfaces. Call these SOs *phases*. Here, Chomsky assumes that Transfer must take place, and that it will take place at specific points in the derivation only (or perhaps, that it can only take place *successfully* at specific points in the derivation). These points, he calls *phases*. In other words, he attempts to derive the notion of *phase* from SMT, rather than from linguistic data. (Although of course linguistic data provided the original

motivation for positing the existence of phases.)

- For minimal computation, we assume that once information has been transferred, it is forgotten, not accessed in further stages of the computation. Thus we formulate a *phase-impenetrability condition* PIC. Note that for narrow syntax, probe into an earlier phase will almost always be blocked by intervention effects. In other words, the lower phase will have a possible but always inactive goal. Perhaps, then, the PIC only holds for the mappings to the interface, its effects following automatically for narrow syntax due to the nature of the system. The edge of the lower phase being excluded, I assume?
- Presumably, the phases correspond to CP and v*P. C is the “left periphery”, possibly involving feature spread from fewer functional heads (possibly only one). (So Chomsky seems to reject an extended left periphery with each feature representing a separate head?) v* is the functional head associated with full argument structure, transitive and experiencer constructions. (Chomsky calls it φ -complete v in *Derivation by Phase*. Other options for v are unaccusative and passive, which are both not φ -complete, because they cannot assign accusative. Hence they are v heads, not v*.) DP may be a phase as well.
- Although it is normally T that is associated with φ -features, there is reason to believe that T’s φ -features are derivative from C; e.g., T manifests these features only when it is selected by C. Presumably, T lacks them in the lexicon, and inherits them from C.
- Agree- and Tense-features are inherited from C, the phase head. If C-T agrees with the goal DP, the latter can remain in-situ under long-distance agree, with all uninterpretable features valued; or it can raise as far as SPEC-T, at which point it is inactivated, with all features valued, and cannot raise further to SPEC-C. We thus derive the A-A’ distinction. I do not believe it is immediately obvious how, at this point. Later on, Chomsky states that an A’-position is defined by the fact that it is attracted by the edge feature of a phase head, typically in SPEC-C or outer SPEC-v*.
- There is ample evidence that the A-A’ distinction exists, and is crucial at the C-I interface. It is therefore expected that it is established in narrow syntax, and inheritance of features of C by T provides a simple mechanism. That is, SMT requires that something that is crucial at the C-I interface be marked in syntax somehow. Since Agree features are related to A-elements and EF with A’-elements, it is natural to separate the two kinds of features and thus establish the A-A’ distinction. When φ -features appear morphologically without tense, they should therefore be regarded as just a morphological effect of agreement, without significance in the syntactic computation. (How is this meant? Does “morphological effect of agreement” mean that there is some form of agreement outside of narrow syntax, not mediated by Agree? I mean, if T cannot be a probe itself, it should not be able to agree unless it is dominated by C.) The inability of TP to be moved or to appear in isolation without C gives further reason to suspect that TP only has phase-like characteristics when selected by C. That is, properties of the “sound side”, as Chomsky formulated it in earlier work, also suggest that TP is not a phase.
- Suppose that EF permits free Merge to the edge, indefinitely. That yields a certain subcategory of recursive systems. (Which subcategory, exactly?) If the EF of an LI is not satisfied, the derivation may crash, but will not always do so. (What determines this?) If EF is minimally satisfied for an LI, then this LI has a complement, to which C-I will assign a certain interpretation, e.g., a theta role. Second EM to v* will also yield a theta-role assignment at C-I.

- It is assumed that only phase heads will trigger operations. As such, IM will only satisfy EF for phase heads. (Thus, raising to SPEC-T must be derivative, via inheritance.)
- Merge applies freely, the resulting SOs being given an interpretation at the interfaces, which can be “deviant”. (What does this mean, exactly? Apparently not exactly “ungrammatical”.) The only empirical requirement is that SM and C-I assign the interpretations that the expression actually has, including many varieties of deviance. A lot can be said here. First, the last sentence seems to imply some form of cartesian dualism: it implies that a structure has an inherent interpretation that is independent from the interpretations given to it by SM and C-I. This would be a fallacy, in my opinion: a structure only has the interpretations given to it by SM and C-I; in an of itself, it means nothing. But perhaps Chomsky merely means that a theory is only correct if it assigns to a structure the interpretation that we empirically find it to have. Second, and this seems more important, Chomsky appears to depart from a strict grammatical vs. ungrammatical dichotomy. Grammaticality is no longer a property of a derivation (i.e., a derivation can no longer be said to “crash”), it is a property of the interpretation of the derivation at C-I and/or SM, both yielding different types of “grammaticality”: C-I yields the more “traditional” kind, while for SM ungrammaticality would be unpronounceability or presumably also incorrect morphology. (E.g., *er laufen* would be interpretable at C-I, but judged deviant at SM because of the incorrect form *laufen*.)
- The label of an SO must be identifiable with minimal search, by some simple algorithm. Here the term *minimal search* is used, which later seems to be a formally defined notion; the formal definition is not given, however. Two options are:
 - In $\{H, \alpha\}$, H an LI, H is the label
 - if α is internally merged to β , forming $\{\alpha, \beta\}$, then the label of β is the label of $\{\alpha, \beta\}$
- These two options suffice for virtually every case. (The exceptions being second EM to v^* , and possibly small clauses.) Sometimes, however, they conflict. One example is the first step of every derivation, which merges two LIs, which may both project, one option perhaps resulting in some form of deviance.
- Another case arises when LI α is internally merged to non-LI β : in this case, both can be the label:

(4) what [C [you wrote t]]

If C projects, this yields an interrogative (sub)clause; but if *what* projects, a free relative object results. So Chomsky exploits the ambiguity inherent in the system, rather than trying to eliminate it. The latter would probably require additional stipulations, so the approach is in line with the general idea. It does make it necessary to show, however, that unwanted derivations can indeed be filtered out elsewhere (at C-I and/or SM). Furthermore, it increases the danger that the model predicts that the computational system is constantly creating derivations that come to nothing, a situation somewhat reminiscent of the days of global, relative economy hell. Not necessarily a desirable one.
- Note that if *what* is the label, C must also be the label, because of feature inheritance by T. Hence, in this case, *both* are the label. Chomsky says both labels “co-exists”, hence are not combined into a single label. Note that we must now also assume that in first merger in every derivation, where two LIs are merged, the resulting structure can have two labels. Again, this means that the model may produce unwanted derivations, which are only filtered out in the C-I module. BT Condition C may be

a similar case. That is, as discussed above, the antecedent (subject) pronoun projects as well as T, the head with which it is merged.

- The conclusion, then, is that the labeling algorithms apply freely, sometimes producing deviant expressions. The outcome will satisfy the empirical conditions on I-language if these are the interpretations actually assigned. In other words, we should not be worried about the labeling algorithms producing results that seem incorrect. If the interpretational systems do indeed flag such structures as incorrect, everything's fine. This seems to confirm the interpretation above that Chomsky no longer defines ungrammaticality as a property of a derivation. (Though it may in fact be the case that he never viewed it as such, that it was just a "misinterpretation" by "the field". Or by me.)
- Identifying internally merged α with its copy/ies but not with other items that have the same feature composition is trivial, if within a phase each selection of an LI from the lexicon is a distinct item, so that all relevant identical items are copies. I do not really see how Chomsky thinks such a formulation is *not* confusing. If one is to identify a copy from a newly selected element, the copy must have some feature identifying it as a copy, which seems to raise questions for the NTC. But since Chomsky mentions in a footnote that remerge is simply the copy theory in its original form, it would seem (as noted earlier) that a copy for Chomsky is not really a separate element, but somehow the same element merged again. Nothing more than phase-level memory is required to identify these properties at the semantic interface C-I, where the information is required. In other words, keeping the two apart does not require much computational load.
- It should be noted that IM can only be eliminated by stipulation, and doing so would furthermore require some other mechanism to distinguish copies from unrelated occurrences. This and other problems mean that eliminating IM has many disadvantages, and it appears to have no compensating advantages.
- The copy theory of movement eliminates the lowering operation required for reconstruction, though how reconstruction works exactly is still an unresolved matter.
- At the phonetic interface, the interpretation of copies faces two conflicting desiderata: (i) ease of processing, which would require copies to be maintained; and (ii) minimization of computation, which would require copies to be deleted. Overwhelmingly, (ii) is correct, which gives additional evidence that language is "designed" so that mapping to C-I approximates SMT, with utility for communication only a secondary factor. Is this really true? Communication requires a two-way mapping between a structural description SD and a phonological representation (that one would be tempted to call PF, but for the history of that term...) which is done by the SM module. The SD→PF mapping is served by (ii), the PF→SD mapping would be served by (i). That means that as soon as one employs IM, one has to make a choice between (i) and (ii), and either way the load for SM increases. The fact that (ii) is chosen may say something about SM, but I do not see how one can conclude from it that the C-I interface is primary. In fact, it does not seem to bear on C-I. It would seem that the option causing the least effort for SM would be to not use IM at all. Recall, however, that IM is primarily used for communicative purposes.
- If minimization of computation is the driving force in spell-out, one would expect that there can be conditions in which some residue of lower copies *is* spelled out. Such cases do indeed seem to exist. Furthermore, *wh in situ* may be accounted for by saying that the lower copy is spelled out.
- For minimal computation, a probe should search the smallest possible domain to find its goal: its c-command domain. Hence there is no m-command, no SPEC-Head relations (except when

the SPEC is itself the probe). The number of specifiers is unlimited (can only be limited by stipulation), the SPEC-COMP distinction reduces to first Merge, second Merge, etc. These assumptions reduce descriptive technology, thus approach SMT, but still have to be shown to be empirically viable. The idea that specifiers can be probes seems to open up the way for Michael Starke's specifier-less syntax, where a SPEC is reanalysed as a "complex" head. (Although I believe that proposal has some open questions with regard to determination of labels.) And it raises the question when a SPEC can be a probe.

- Minimal search is not uniquely defined in XP-YP structures where neither XP nor YP is a head: the "wrong choice" yields island effects. This remark is rather unclear to me. It appears that minimal search is (or should be) a strictly defined notion, and no doubt Chomsky has a specific idea in mind, but he does not seem to make it explicit. The c-command domain for a probe is clear: its sister and everything in it. This remark, however, appears to suggest that certain parts of the c-command domain of a probe are excluded from minimal search. After all, in a node (XP,YP), if XP is a probe, it is pretty clear what the search domain is: YP, down to the next phase. Hence when Chomsky says "where neither XP nor YP is a head", he appears to mean "where neither XP nor YP is a head *or a probe*". So he is talking about delimiting the search domain for a probe P that is outside XP and YP. That seems indeed the case in the subject island examples he discusses below.
- Consider Huang's CED (Condition on Extraction Domain) effects; (i.e., the ban on extraction from subjects and adjuncts). Consider the subject island subcase (adjuncts would follow under a set-Merge analysis of adjuncts). There is reason to assume that it is not the surface subject position that is the island:

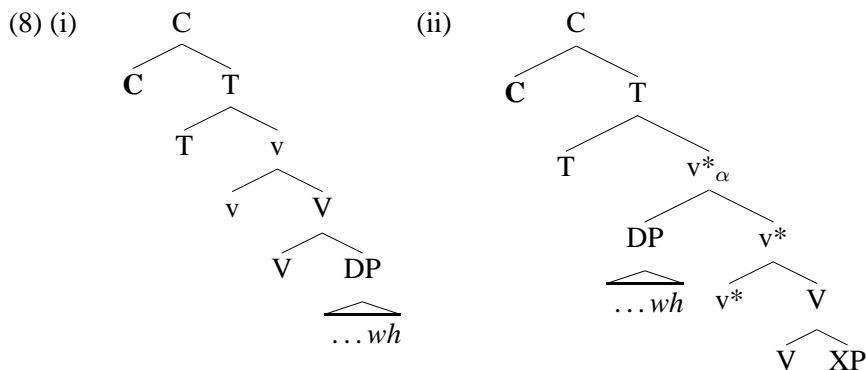
- (5) (i) it was the CAR (not the TRUCK) of which [they found the (driver, picture)]
(ii) of which car did [they find the (driver, picture)]
(6) (i) *it was the CAR (not the TRUCK) of which [the (driver, picture) caused a scandal]
(ii) *of which car did [the (driver, picture) cause a scandal]?

(5) and (6) are standard extraction cases, from object (licit) and subject (illicit) position, respectively. Compare with the following examples:

- (7) (i) it was the CAR (not the TRUCK) of which [the (driver, picture) was found]
(ii) of which car was [the (driver, picture) awarded a prize]?

(7) pairs with (5), not with (6). Although the DP from which the *wh*-element is extracted (*the (driver, picture) of which*) is in surface subject position, extraction is possible.

The relevant base structures are the following:



In tree (8i), v is unaccusative/passive, hence not a v^* , hence not a phase. So only (8ii) has the internal phase α , headed by v^* . We now have the right distinction, though it remains to be explained. The probe we are talking about is C (bold in the trees above). Since v in (8i) is not a phase, C can probe into it. In (8ii), v^* is a phase, but the goal wh is in a DP that is in the edge of v^* , so the phase boundary doesn't make it inaccessible. Now, the (XP,YP) node Chomsky talks about appears to be (DP, v^*), as neither of these elements is a head. So in this node, minimal search for the probe C is not defined. This appears to suggest that the search domain is not just the c -command domain, but rather a path down the sister node, in which at each node only one subnode can be selected. The "wrong choice" (i.e., DP in the current example) leads to island effects. What makes a wrong choice the wrong choice, however, is not clear. Note that the non-phase status of v in (8i) is relevant: if it were a phase, DP would have to be raised to SPEC- v before C-T were merged, meaning a configuration similar to (8ii) would arise. It seems to me that the intuition Chomsky wants to express here is that there is only one particular configuration in which minimal search is not defined, and that is when two non-heads are merged. Why minimal search is not defined there remains to be explained.

- There are further consequences. If T were to raise the DP in (7) before the wh -element were probed by C , a configuration similar to the one in tree (8ii) would ensue, meaning an island violation would be expected in (7). Therefore, both the A- and the A'-movement operations are triggered by probes in C , and the two operations proceed in parallel. I don't recall any real precursor to the idea that operations can operate in parallel, and it seems rather a revolutionary one. One wonders if it might possibly be a solution to the problem for the extension condition that is inherent in the idea that movement to SPEC-T is triggered by C . It only could, however, if EM of C also operates in parallel to the two probe/IM operations. And that seems to take one awfully close to a representational model.
- It remains to explain why the probe for wh -movement cannot access the wh -phrase within the external argument of the phase α . That could reduce to a locality condition: *which* in α is embedded in the lower phase, which has already been passed in the derivation. To put it differently: when the probe tries to match *which* inside α it has to pass two phases: the v^* phase, and the DP phase. Barriers, anyone?
- The same would of course also be true for the extraction in (5), because V in (5) has a full argument structure, and hence is dominated by a v^* , just like (6). The extraction of wh out of the object is possible in (5) because wh first moves to SPEC, v^* , with v^* acting as probe. Extraction out of the subject in (6) is impossible because v^* cannot act as a probe for something in its own specifier. For this reason, SPEC-to-SPEC movement is always impossible. That's to say, within a phase. Movement from SPEC- v^* to SPEC-T or SPEC-C is possible. One question is raised here: if DP is also a phase, it will have an EF, and hence should be able to probe for the wh in its complement. If probe from the higher C is able to reach the SPEC of SPEC- v^* , it could still reach the wh -phrase, in spite of the fact that the containing DP is in subject position.
- This reinforces the conclusion that C has two probes: EF and an Agree feature (φ -features). EF raises wh to C , φ -features raise the DP to T. This raises the usual two questions: how and why?
- The *how* has already been suggested for other reasons: T inherits its Agree-features from C , and then derivatively serves as a probe at the phase level CP. The motivation may trace back to a C-I interface requirement that both argument and operator-variable structures be available, analogous to the requirement of semantic duality that is satisfied in an optimal way by the A-A' distinction.

- Ideally, transmission of Agree-features should be a property of phase heads in general, not just of C. Hence v^* should transmit its Agree-feature to V and probe of an object with structural Case by v^* should be able to raise it to SPEC-V. (Strangely enough, no longer to SPEC- v^* .) Such facts have been known, and now it becomes clear why they should exist: they are properties of phase heads in general. Furthermore, if the suggestions above about motivation prove to be accurate, the curious phenomenon of raising to SPEC-V follows from the C-I requirement that the A-A'-distinction must be observed at the CP-phase, supplemented by third-factor conditions on efficient computation. This remark is unclear to me. Does Chomsky mean that C has the property that it transmits its Agree-features to the LI it selects because of the A-A'-distinction (as he assumes earlier), and that because all phase heads should be sufficiently similar, v^* also has this property?
- Another question is whether inheritance is obligatory or optional. For C-T, that raises familiar questions about universality of EPP and about mechanisms of agreement. For v^* -V, properties of Binding Theory Condition B indicate that the rule must be obligatory, by the general logic of the clause-mate principles of Postal-Lasnik-Saito. Thus *him* in the next example is necessarily free:

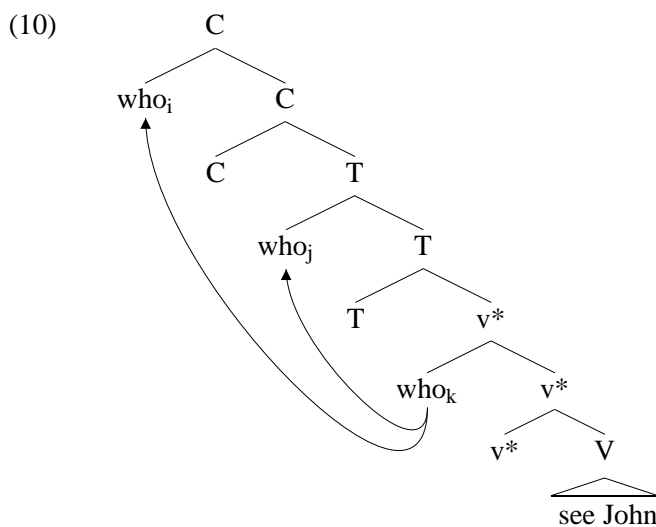
the slave expected [(the picture, the owner) of him to be somewhere else]

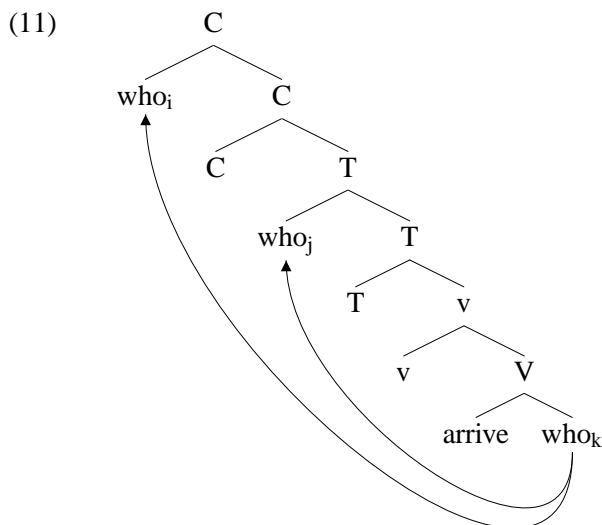
What is meant here, I assume, is that according to Postal a.o. the embedded subject in an ECM construction moves up to the higher clause, which Chomsky here attributes to movement to SPEC-V, after attraction by v^* -V. The fact that *him* in this example is free, indicates that the movement must take place: the movement puts *(the picture, the owner) of him* in a binding domain with the matrix subject *the slave* (since the DP has no subject of itself) and hence co-reference of the subject and the pronoun is no longer licensed. The *for-to* analogue seems to allow the bound option more readily, as would be expected. That is:

the slave expected [for (the picture, the owner) of him to be somewhere else]

In this case, the embedded subject does not move up to the higher clause, and hence a bound reading would be expected to be possible. Lasnik, however, has given arguments to the contrary, so both for C-T and for v^* -V the matter is unsettled.

- What is true for (5)-(7), should be true for *wh*-questions in general. Consider:





In (10), in the v^* -phase, agreement between v^* and *John* (not indicated in the tree) values all interpretable features. In the C-phase, both the EF and Agree-features of C seek *who*, which as a result is raised twice, once to SPEC-T and once to SPEC-C. That means there is a relation between who_i and who_k , and one between who_j and who_k , but *not* between who_i and who_j . There are two A-chains: (who_j, who_k) and (who_k) , each an argument (of what?), with who_i the operator ranging over the A-chains, interpreted as bound variables. (So who_i does bind who_j , even though there is no direct movement relation)

- Although both A-chains are invisible, familiar properties of A-movement (binding, scope, weak cross-over, etc.) show that there really is a copy of *who* in SPEC-T. So we have standard oppositions such as:

(12)(a) *who* was never seen, **who* was there never seen

Here, there is an expletive *there* in SPEC-T, while in order for the derivation to converge, a copy of *who* should move to SPEC-T, as seen in (10)/(11). Note that the theory allows multiple specifiers, so in principle there should be nothing against two SPEC-T positions, one containing the expletive and one containing the relevant copy of *who*. The problem is therefore not that the position is occupied by the expletive, but that the Agree features of T are satisfied by *there*, so that *who* cannot be attracted and its Case feature remains unvalued.

(12)(b) who_i seems to his_i friends to be preferable, * who_i do you seem to his_i friends to prefer

This is a weak crossover case. It's generally the case that a pronoun can be coreferential with a c-commanding *wh*-element when the *wh* originates in subject position, but not when it originates in object position:

who_i [_{TP} ~~*who*~~ likes his_i mother], * who_i does his_i mother [_{VP} like ~~*who*~~]

(12b) makes an important point: it is not the fact that the *wh*-element moves over the co-indexed pronoun that causes ungrammaticality, since in the first clause of (12b) *who* also moves over *his* (the base position being SPEC of *to be* or (rather) SPEC of *preferable*). The observation is that when SPEC-T is occupied by a copy of the *wh*-element, the coindexation with a c-commanded pronoun is ok. When there is no copy of the *wh*-element in SPEC-T, coindexation with a c-commanded pronoun leads to ungrammaticality. Therefore, the contrast in (12b) shows that there is indeed movement to SPEC-T, even when the *wh*-element comes from a lower subject position.

- It has been conventionally assumed that in cases such as (10)/(11) there was a non-uniform A'-A chain. Now we see, however, that there are in fact two chains (three, if one counts the chain (who_k)), a uniform A chain and a non-uniform A'-A chain.
- In principle, (10)/(11) could be formed in another way: EF of C could, instead of probing for *who_k* simultaneously with T, probe for *who_j*, but then not simultaneously with but *after* T. This looks suspicious, however, because we know that EF of C cannot extract the PP complement from within SPEC-T (as that would constitute a standard subject island violation). It must be, then, that SPEC-T is impenetrable to EF, or more naturally, invisible to EF. So an A-chain becomes invisible to further computation when its uninterpretable features are valued. That principle incorporates the effects of the earlier inactivity condition.
- For our purposes, we can define an A'-position as one that is attracted by an edge-feature of a phase head, hence typically SPEC-C and outer SPEC-v*. (Note: only *outer* SPEC-v*, not the inner SPEC-v*, because that is the position where the subject is (externally) merged.) Other positions are A-positions. That is, A- and A'-positions are no longer distinguished by their structural positions, but by the manner in which they are derived. As a result, successive-cyclic A'-movement creates a uniform A'-chain (with only the lower position a possible A-position). Intermediate positions do not induce binding effects or have other A-position properties, whatever their structural status.
- With all operations driven by the phase head, the only A-chains are completed A-chains with all features valued, either inherently or by Agree. (Which implies that A'-movement will not take place from the head of an A-chain.) It follows that lower copies ("traces") are invisible. I assume because the entire chain is invisible, due to the fact that all its features are valued.
- Suppose PH is a phase head selecting PH_s. (That is, PH=C selecting PH_s=T, or PH=v* selecting PH_s=V.) Then PH can raise an XP to SPEC-PH, but only from its base position. If XP has moved to SPEC-PH_s, (due to Agree features on PH_s inherited from PH), then XP is invisible: it cannot be raised and nothing can be extracted from it.
- If, however, EF of PH raises XP to SPEC-PH, it (by definition) no longer heads an A-chain, and is subject to raising or extraction by a higher EF. I do not know how to interpret "subject to extraction" here. We have just seen that nothing can be extracted from an XP in SPEC-v*, and the same should be true for SPEC-C. Chomsky seems to think so too, given his next remark: Extraction from this A'-position should be on a par with extraction from an external argument. Thus (13) should have about the same status as the subject-island violations (6):

(13) of which car did you wonder [which [(picture, driver)] [caused a scandal]]

This would be expected because *which* (*picture, driver*), from which *which car* has been extracted, is in SPEC-C of the embedded clause. I do not know why Chomsky says "should have" and not "has".
- Raising by EF of an XP cannot follow long distance Agree, because that would make XP invisible. Furthermore, XP could not be raised before agreement, or its Case feature will be unvalued. (Note: raising here is to SPEC-C, while the Case feature would be valued by Agree between T and XP. So if raising were to apply before Agree, the element raised to SPEC-C (and most likely the element later to be spelled out) will not have a valued Case feature. In earlier models, such considerations never played a role, because raising to SPEC-C would proceed from SPEC-T, where Case has already been valued. Now that raising to SPEC-C proceeds from the base position, valuation of Case is not automatically given.) It follows that the edge and Agree features of PH operate in parallel.

- Essentially, all options are open: the edge and Agree features can apply in either order, or simultaneously, with only some options converging. Again Chomsky implies that the computational system creates structures more or less at random, which are then filtered by the interpretational modules C-I and SM; i.e., the computational system *overgenerates*.
- What holds for *wh*-movement should hold for *A'*-movement in general. Suppose that the edge feature of the phase head is indiscriminate: it can seek any goal in its domain, with restrictions determined by other factors. We adopt Rizzi's approach to the left periphery: what is raised is identified as a topic by the final position it reaches, and any extra specification (i.e., Topic features, *wh*-features, etc.) is redundant. We need not postulate an uninterpretable feature that induces movement, and can thus overcome a long-standing problem about crash at the lower phase-levels in successive-cyclic movement. Again a rather revolutionary idea. It has always been assumed that an element that moves to the C-domain is attracted by a matching feature: a *wh*-element moves to SPEC-C because C has an uninterpretable [+wh] feature. This has always been a bit of a problem: the *wh*-feature on C has no function other than attracting a *wh*-element. It cannot be argued that it marks the clause as interrogative, because it is uninterpretable. It cannot be argued to be interpretable, because that would mean the *wh*-element doesn't have to move to check/value it. The same reasoning applies to any other type of element that moves to SPEC-C, and there is a similar problem for elements raised out of lower clauses: what is the feature that makes the element move to the embedded SPEC-C? Now Chomsky says that EF can attract anything in its search domain, as long as it is not blocked for some other reason, and presumably means that it is up to the interpretational modules to assign meaning to the resulting derivations, or reject them. Again, this implies that the computational system overgenerates.
- Further elaboration depends on how the relevant structures are to be analyzed properly. Chomsky then gives a few possibilities with respect to *wh*-phrases: they may or may not have an interpretable *wh*-feature, and they may or may not move to a position where they can be interpreted. Any such structure will converge, but may under circumstances be uninterpretable.
- Note that there should be no superiority effect for multiple *wh*-phrases: any can be targeted for movement. Other reasons lead to the same conclusion:

(14) C [T [who [v* [see what]]]]

At the lower phase v*, the subject *who* does not intervene (EF of v* cannot probe into SPEC-v*, as we have seen above), so *what* can be raised to the edge. So in principle one would expect that *what* can be moved to SPEC of matrix C, with the subject *wh who* still in situ. It remains to explain superiority effects in the languages that show them, though this does not seem easy. The remarks here raise a few questions. In principle EF of the phase head can attract any element in its search domain, how then do we account for a couple of common distinctions? Why is it that one language has obligatory *wh ex situ*, while others have obligatory *wh in situ*? Why in multiple *wh*-questions do some languages have obligatory movement of all *wh*-elements, while others only move one? It would appear that the computational system can derive all possibilities, and has no way whatsoever to filter out the ones that a particular language does not accept. It seems that they can now only be filtered out at the interfaces, which means that either C-I or SM must do it. C-I cannot really be responsible for it, because it would mean that the C-I module is language-dependent. We are therefore left with the SM module. In itself, the SM module must also be universal, but the material it works with, the phonological forms of lexical elements, are not. So we may have eliminated a bunch of ugly, uninterpretable features at the C head in syntax, but now we get them back as properties of lexical items that SM is sensitive to: Japanese must have some feature on *wh*-elements that says that the lower copy must be spelled out,

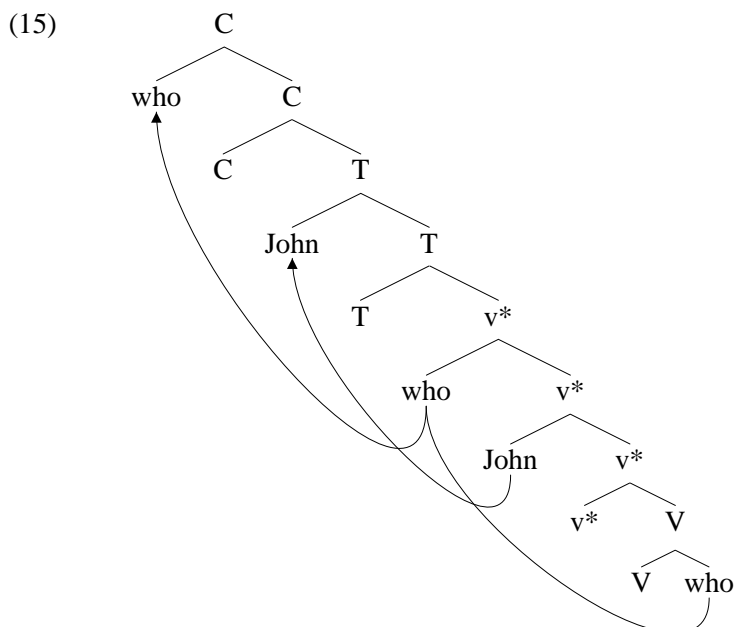
while in English they have a feature mandating spell-out of the higher copy. (Either that, or we assume that spell-out of the higher copy is default.)

Alternatively, we must find some way to say that in Japanese only derivations in which *wh*-elements do not move are acceptable at the interfaces. I don't really see how we can do this.

Wh in situ in English, i.e., to form echo questions, could be accounted for if we say that these are structures in which *wh* hasn't moved, which results in a different interpretation. But multiple *wh*-questions aren't so easy, it seems.

What one would like to say, intuitively, is that the C-I module judges the structures not just on the basis of the SO that it receives, but on the combination of the SO and the PF built from the SO at SM. The model does not allow for that option, however.

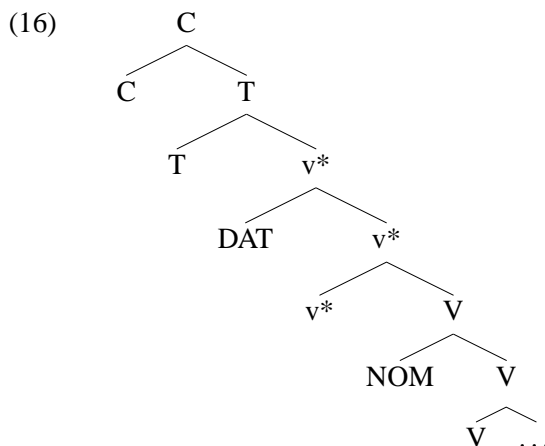
- A'-chains function as A-chains with regard to intervention: if uniform, only the full chain (equivalently, its head) is the object that intervenes. We can see this in the following structure:



John here crosses the lower copy of *who* in the chain (SPEC-C, OUTER-SPEC- v^*), but it does not cross the entire chain, so the A'-chain cannot cause intervention effects for *John*.

Note that there is a problem with assuming that *who* moves to outer SPEC- v^* , not inner SPEC- v^* . After all, we assumed tucking in, and *John* is externally merged in SPEC- v^* , hence we can assume that it takes place before internal merge of *who* in SPEC- v^* . The reason for assuming the object moves to outer SPEC- v^* is, I assume, that it is an A'-position (being triggered by EF of v^*), and it would seem odd to say that an A'-position is below an A-position (the position where the subject is externally merged). However, if third merge can target both the inner and the outer SPEC, we would need a mechanism to determine which position must be targeted in which circumstances. Either that, or allow overgeneration of such structures.

- Another illustration has been discovered in Icelandic. Consider the dative-nominative experiencer construction (16) (e.g., “to-someone_{DAT} seem [the horses_{NOM} are slow]”):

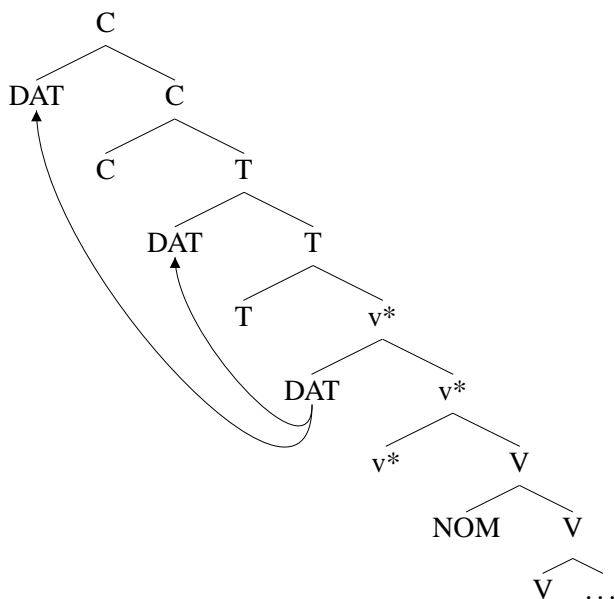


If DAT remains in situ, in an expletive constructions (it seems to-someone [the horses are slow]) it blocks T-NOM agreement, as expected, because DAT intervenes. If DAT is raised to SPEC-T, T-NOM agreement is permitted, again as expected: there is no intervening argument, because only the lower copy in the chain (SPEC-T, SPEC-v*) intervenes. But if DAT is *wh*-moved, it blocks agreement, which is paradoxical, since it appears to be the lower copy of an A-chain. That is, in the older model, DAT would first move to SPEC-T and from there to SPEC-C. As such, one would expect T-NOM agreement to be possible, because only the lower copy of DAT intervenes, which does not block agreement, as seen in the case where DAT moves to SPEC-T. The suggested solution is that *wh*-movement of DAT goes to SPEC-C directly, not through SPEC-T. As such, there is an A-chain consisting of *only* SPEC-v*, which *does* block T-NOM agreement, as seen in the case where DAT remains in SPEC-v* (the expletive case).

- That seems basically right, but causes other problems, because we do want to have the A-chain (SPEC-T, SPEC-v*), for reasons already mentioned. The effects mentioned in (12), that is. The desired result follows if both operations, A- and A'-movement, are driven by the phase head C. That will leave no relation between SPEC-C and SPEC-T, but an operator-argument relation between SPEC-C and each of the two A-chains, (SPEC-T, SPEC-v*) (formed by the Agree-feature of C) and (SPEC-v*) (formed by the edge feature of C).

This is strange. Of course, the edge feature of C also forms an A'-chain (SPEC-C, SPEC-v*), but that is irrelevant here. What Chomsky suggests is that EF-driven movement not only creates the familiar A'-chain, but also an additional A-chain consisting of only the source location of the A'-movement.

What happens is this: the phrase we are talking about is something along the lines of "to-whom_{DAT} seems [the horses_{NOM} are slow]". Here, agreement between T and NOM (*seems* and *the horses*) is ruled out.

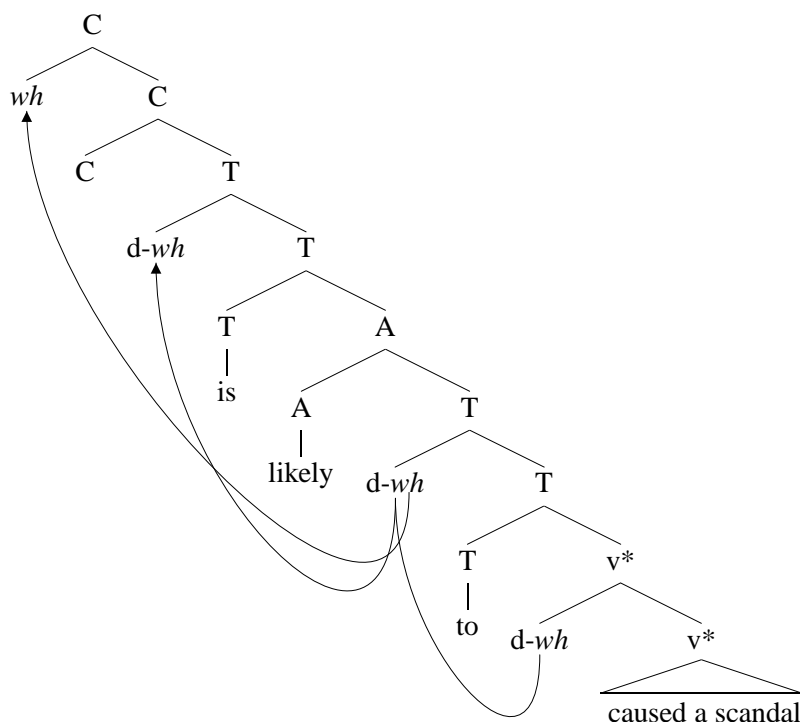


The movement from SPEC- v^* to SPEC-T creates an A-chain. If it were only for this A-chain, agreement between T-NOM would no longer be ruled out, because only the lower copy in the chain intervenes, not the entire chain, as we've seen above for the case "to-someone seem [the horses are slow]". But in this case, where the dative is a *wh*-element and has moved to SPEC-C, agreement between T-NOM is not permitted. Chomsky accounts for this by saying that apart from the A-chain (SPEC-T, SPEC- v^*) created by T-DAT agreement (!), another A'-chain is created, namely (SPEC- v^*), by the A'-movement (!) of DAT to SPEC-C.

The two exclamation points mark two questions: first, Chomsky claims that there is a unitary A-chain that is *created* by A'-movement. A rather surprising move, if I may say so. Second, Chomsky says that movement of DAT to SPEC-T results from the Agree features of C inherited by T. And exactly that movement is the one that allows T-NOM agreement. But if T's (inherited) Agree features have already been valued, the question is raised why there is any need for further T-NOM agreement. Apparently, DAT cannot value all of T's features, but if so, the question still remains why T cannot just take a default form. After all, when DAT does not move, default features on T *are* possible.

- Looking at successive-cyclic A-movement, we reach similar conclusions. Take for example (18) (with *t* indicating the lower copies of *of which* in A-positions):

- (18) (i) it is the CAR (not the TRUCK) of which [_{TP} the (driver, picture) is likely [_{TP} *t* to [_{V*P} *t* cause a scandal]]]
- (ii) of which car is [_{TP} the (driver, picture) likely [_{TP} *t* to [_{V*P} *t* cause a scandal]]]



- Although extraction is from the subject in (18), it is grammatical, unlike its counterpart without long extraction in (6). This follows if successive-cyclic movement must proceed through SPEC-T (due to the still mysterious EPP condition). When the subject lands in SPEC-T of the infinitive, EF of the matrix C extracts the *wh*-element. No deep search is needed, because no phase boundaries are crossed. I.e., the *to*-infinitive is a TP, not a CP. The two operations interweave. (That is, A'-movement proceeds before the relevant element has reached its highest A-position.)
- The same conclusions hold for ECM:
(19) of which car did they believe the (driver, picture) to have caused a scandal
The *wh*-phrase of which car must have raised from an intermediate position, SPEC-T of the ECM infinitival, before it reaches SPEC-V.
- We see, then, that the two searches driven by the phase head (probe by EF and probe by Agree-features) operate in parallel, and can even interweave. What yields the subject-island effect, it appears, is search that goes too deeply into a phase already passed, not the difference between base and surface position.
- The Agree feature of C raises XP to SPEC-T (by inheritance), while its edge feature raises XP (or part of it) to SPEC-C.
- The generalized inactivity condition bars extraction from matrix SPEC-T (and if fully generalized, raising of full XP from that position). (That is, an XP raised to SPEC-T has all its features valued. Essentially, A'-movement can never originate from SPEC-T.)
- Such constructions as (19) provide an independent reason for ECM-raising to SPEC-V. (Note that before, an ECM subject, and indeed objects in general, would raise to SPEC-v*. It has now become

SPEC-V on analogy with C-T raising the subject to SPEC-T, not SPEC-C.) And extraction of complement provides an independent reason, alongside of binding and (perhaps) reconstruction effects, for successive-cyclic A-movement through SPEC-T. (Which, as just remarked, is essentially still unexplained.)

- Note that extraposition of PP-complements is sharply different from extraposition:

(20) *the (driver, picture) is likely to cause a scandal of the car

(21) *the (driver, picture) caused a scandal of the car

We conclude, then, that despite some superficial similarities, extraposition and A'-movement of PP-complement are entirely different phenomena. If they weren't, (20) would be expected to be much better than (21). PP-extraposition is part of the mapping to the SM interface, hence part of SPELL-OUT. If so, it should be restricted to the interior of a phase, thus allowing extraposition from object but not from EA subject. An assumption that might be problematic, if DPs are phases.

- The size of phases is in part (and in what part not?) determined by uninterpretable features (uF). The values of uF are redundant, fixed in the course of the derivation. Because of their redundancy, we expect uF to be unvalued in the lexicon, and to be deleted before the derivation reaches C-I, i.e., before or during Transfer. Since they may be expressed phonologically, they cannot be deleted before Transfer to SM, and they must be valued when Transfer takes place. Furthermore, the stage where uF are transferred must be the same stage for both Transfer operations, because once uF are valued, they are indistinguishable from the interpretable features that valued them. And could therefore lead to crashes, I assume. Once valued, uF are deleted by the mapping to C-I, and given whatever phonetic properties they have in particular I-languages by SM.
- By definition, the stages where Transfer takes place are phases. Since uF must be valued when Transfer takes place, we find further support that v*P and CP are phases, because it is at those stages that uF are valued. Of course, for subjects this is not really true, which is why Chomsky follows this remark with: For subject, the conclusion is based on the assumption that TP is not a phase, for reasons discussed, so that T operates as a probe only derivatively by virtue of its relation to C. Isn't there some circularity here? TP is not a phase, and we want the phase heads to be responsible for movement. So we devise a mechanism (inheritance of Agree features) in which it is C, not T, that is ultimately responsible for agreement on T. Then we argue that uF must be valued when transferred, and see: this happens only at the C level, not at the T level! So we conclude that C is the phase, not T. But that is the assumption we started out with.
- There is also morphological evidence that CP and v*P are the phases. Just for these two categories the edge is sometimes morphologically marked in successive-cyclic movement, (for complementizers, this is true, but what marking exists for v*?) with the effect of movement through SPEC-C sometimes found in the subject-agreement domain, another reason to suspect that T-agreement is derivative from properties of C. Chomsky (conveniently) leaves out any references to the "effects of movement through SPEC-C sometimes found in the subject-agreement domain", so it is not clear what he is talking about. If he means complementizer agreement, then one faces the serious question why complementizer agreement is dependent on the presence of T-agreement (Eric Fuß's dissertation) rather than the other way around.
- Phases should, presumably, be as small as possible, to minimize computation after Transfer and to capture as fully as possibly the cyclic/compositional character of mappings to the interface.

There are two basic cases to consider:

- (I) SO cannot be transferred to the SM interface (“spelled out”) if it is subsequently going to move
- (II) SO cannot be moved to an edge unless it can be spelled out right there, satisfying any uninterpretable features by long-distance Agree

(I) is transparent, unless more complex apparatus is introduced that we would hope to avoid. (It's quite unclear to me what Chomsky wants to say here.)

(II) has to be sharpened. It conforms to a fairly general empirical observation that should be captured:

- (III) In a probe-goal relation, the goal can be spelled out only in-situ (under long-distance Agree) or at the probe (under internal Merge)

Again, I do not see where Chomsky wants to go with this. (II) seems to be contradictory with (III): (III) says that an element can only be spelled out in its base position or in the highest position. (II) seems to say that an element can only move if it can be spelled out in the position where it moves to, which in successive-cyclic movement includes positions that (III) would exclude as locus of spell-out.

- There is much evidence that indicate that movement takes place in local steps, with the moving element passing through intermediate positions in which it cannot stop.
- There are differences, however, between A'- and A-movement with regard to local steps. These remain to be explained.
- A consequence of the conclusion that the Agree-feature belongs to C, and to T only derivatively, is that it is in the same region as the left-periphery head for Focus. This would support the idea that agreement and Focus are two values of the same parameter, with some languages making φ -features more prominent (English), while others make Focus more prominent (Japanese).
- Let us turn finally to the mysterious property EPP, which is considerably more general than the problem for which it was originally formulated (the necessary filling of SPEC-T in English). E.g., v typically permits both long-distance agreement and raising, while v* does not:

(22) *there will [a student [v* take the class]]

The point here being that (22) only leads to a grammatical derivation when the subject is raised:

(22a) a student will [~~a student~~ [v* take the class]]

However, if the lower verb is unaccusative (i.e., has a vP, not a v*P), both raising and LD-agree are possible:

(22b) there will [v arrive a student] (LD-agree)

(22c) a student will [v arrive ~~a student~~] (raising)

I'm not entirely sure how this problem is related to the EPP. The EPP seems present in all constructions, it's just that with v it can be satisfied by an expletive, while with v* it cannot. Does Chomsky want to say the EPP comes in different flavours?

- For infinitivals, we may be able to disregard control structures: subject is null so its structural position is uncertain, and they presumably fall under CP structures in any event. That is, we cannot see if EPP exists here, but if it does, it should find the same explanation as in main clauses. That leaves raising/ECM infinitivals (which Chomsky assumes are TPs, not CPs) and tensed clauses, and the analogous v*-V issues, all of which remain unexplained.
- It is tempting to ask whether EPP can be reformulated in terms of feature inheritance. Perhaps T inherits from C not only Agree but also EF, and by some kind of feature spread, this extends to all Ts in the phase. (Rather a mysterious move, I'd say.) Operations then proceed as before. If there is no accessible NOM, T will have default or null morphology. If nothing is raised, the inherited EF can be satisfied by EM, necessarily of an expletive, because no argument role is available. Personally, I'm not convinced this move will work at all. It doesn't seem really feasible to say that EPP is the result of T inheriting EF from C, because raising/ECM infinitives are also subject to EPP. In other words, trying to reduce EPP to an EF would amount to saying that T *can* function as a probe independently from C.