The Linear Correspondence Axiom and the "Standard" X'-Schema*

Yafei Li
University of Wisconsin

ABSTRACT

Contrary to the general interpretation that Kayne's (1994) theory of the X'-structure doesn't permit intermediate projections (i.e., X'), such projections are actually allowed in the theory, but only when they don't have any sister node. In other words, X' is possible only if it serves no structural function whatsoever. The way to avoid this strange consequence is to exclude X' node from being c-commanders. In addition to keeping the desirable features of Kayne's theory, this modification has two related consequences. (1) It provides a way to represent the specifier and the adjuncts in a structurally different manner, a distinction with empirical support. (2) The resulting structure is almost isomorphic to the "standard" X'-theory, in contrast to Kayne's rather substantial departure from it. A more radical extension of this line of thinking is also explored, with both desirable implications on X°-movement and open issues to be worked out.

Key Words: Linear Correspondence Axiom, X'-theory, X°-movement

Kayne (1994) argues that several fundamental assumptions of X'-theory follow from his Linear Correspondence Axiom (LCA). These consequences of the LCA fall into two groups. The first group, including binary branching, the single head for each phrase, and the phrasal nature of the complement, are derived under the "standard" X'-structure as proposed in Chomsky (1986) and many subsequent works. The second group, containing the unique Spec position and an obligatory head for each phrase, rely in their derivations on a structure that drastically departs from the standard X'-schema. In section 1, I show that this second part of Kayne's theory also yields a structure that is linguistically

* I am grateful to two anonymous reviewers' generous comments on this short article. As the reader can see, the final version has benefited greatly from their insights.

strange. In section 2, I argue that this strange structure can still be interpreted within Kayne's theory, if we adopt a proposal in Chomsky (1995, 1996). However, the resulting theory pretty much restores the standard X'-schema, in spirit if not in form. In the appendix, I briefly discuss a more radical extension of my solution, suggested by a reviewer.

1. The LCA and its unexpected consequence

Kayne's LCA and related definitions are given below:

(1) X c-commands Y iff X and Y are full nodes, X excludes Y, and every full node dominating X dominates Y.

(2) Let X, Y, . . . be nonterminals in a phrase marker, $\langle X, Y \rangle$ be an ordered pair in which X c-commands Y but Y doesn't c-command X (asymmetric c-command), and d be the dominance relation from nonterminals to terminals; then
   a. d(X)=the set of all t such that t=a terminal which X dominates.
   b. d $\langle X, Y \rangle$=the set of all $\langle x, y \rangle$ such that $x \in d(X)$ and $y \in d(Y)$, where $\langle x, y \rangle$ indicates that x temporally precedes y.
   c. d(A)=the union of all d $\langle Xi, Yi \rangle$, for $0 < i \leq n$, where A is the set of all $\langle X, Y \rangle$ in a given tree, with n=the size of A.

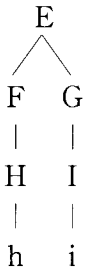
(3) Linear Correspondence Axiom
   Let T be the set of all terminals in a given tree and A be as defined above, then
   d(A)=a total ordering of T.[1]

Mathematical notations aside, the LCA is aimed to capture the simple idea that there is a direct correspondence between asymmetric c-command and temporal precedence: If a nonterminal node X asymmetrically c-commands another nonterminal node Y, then every terminal dominated by X necessarily precedes every terminal dominated by Y; conversely, if a terminal x temporally precedes a terminal y, then there must be a nonterminal node dominating x that asymmetrically c-commands a nonterminal node dominating y. Illustrations of how (1)-(3) work are provided in the course of discussion below.

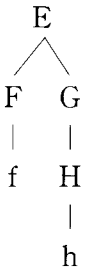
1. In set theory, a total order, also called a linear order, is a binary relation that is either transitive, irreflexive, asymmetric, and connected (i.e., a connected strong order) or transitive, reflexive, antisymmetric, and connected (i.e., a connected weak order). In reality, Kayne's LCA is only about strong orders.

Kayne's derivation of an obligatory head in each phrase is illustrated in (4):

(4) a.

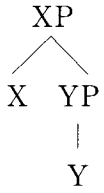


b.



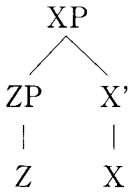
In (4a), the nonterminal E is a phrase containing two other phrasal nodes F and G, with H and I being their respective heads and h and i the terminals they each dominate. In this tree, T (i.e., the set of terminals)={h, i} and A (i.e., the set of ordered pairs of nonterminals with respect to asymmetric c-command)={⟨F, I⟩, ⟨G, H⟩}, so d(A)={⟨h, i⟩, ⟨i, h⟩}≠asymmetric. Since d(A) is not a total ordering of T, (4a) is not a well-formed tree by the LCA. In (4b), T={f, h} and A={⟨F, H⟩}, with d(A)={⟨f, h⟩}=a total ordering of T. Thus, the tree is permitted by the LCA. The crucial difference between the two structures is that F in (4b) is a head because it immediately dominates a terminal whereas F in (4a) is not a head. It follows that the phrase E must have a head. Converted to the X'-notation, (4b) is represented in (5):

(5)



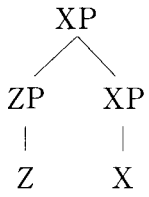
Since a phrase in this theory is defined as any nonterminal immediately dominating its head, the intermediate X' node in the standard X'-schema is disallowed by the LCA because it is also a phrasal node:

(6)



Put differently, (6) is ruled out because XP corresponds to E in (4a), ZP corresponds to F, and X' corresponds to the phrasal node G. In order to allow a Spec position in a phrase, Kayne proposes to replace X' with a segment of XP:

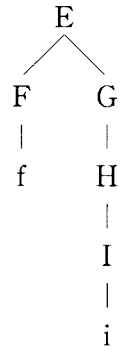
(7)



Since c-command is defined as a relation between full nodes (cf. def. (1)), the lower XP segment doesn't c-command Z, in contrast to the X' node in (6) which as a full node does c-command Z. Neither does the full node XP c-command Z because it doesn't exclude Z. It follows that ZP asymmetrically c-commands X but Z is not asymmetrically c-commanded by XP. In other words, $\langle ZP, X\rangle \in A$ but $\langle XP, Z\rangle \notin A$. Assuming that ZP ultimately dominates the terminal z and X the terminal x, this means that $\langle z, x\rangle \in d(A)$ while $\langle x, z\rangle \notin d(A)$, a result permitted by the LCA. Therefore, Kayne concludes that the specifier must be adjoined to a phrase. Note that the intermediate X' node in the standard X'-theory is not allowed in such a structure.

But consider the following tree, in which $T=\{f, i\}$ and $A=\{\langle F, H\rangle, \langle F, I\rangle\}$:

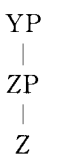
(8)



Because both $\langle F, H\rangle$ and $\langle F, I\rangle$ map to $\langle f, i\rangle$ under d as defined in (2), $d(A)=\{\langle f, i\rangle\}$ = a total ordering of T (trivially). So (8) is well-formed. However, since H doesn't immediately dominate a terminal, it is a phrasal node. So (8) can be converted to (9):[2]

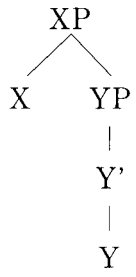
2. Strictly speaking, the subtree G also corresponds to the X'-structure in (i), among others:

(i)



But such a structure can be excluded in Kayne's theory with whatever mechanism forces G in (4b) to be a projection of H instead of a headless phrase.

(9)



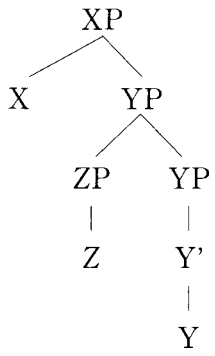
In other words, X' is still allowed, provided that there is no specifier in the phrase.

This conclusion seems strange. From the linguistic point of view, a basic reason for the X' node is to have a subphrase containing everything in XP but the specifier. Now Kayne's theory predicts that X' is possible only when the specifier is absent. Why should this intermediate node be permitted only when it is absolutely useless? How do we interpret this unexpected result of Kayne's theory?

2. A Solution

Consider two more consequences of the LCA that Kayne does not discuss. First, the phrase-adjoined specifier and the "useless" X' node can co-occur in the same phrase. In (10), YP is the complement of X and ZP is the specifier in YP.[3]

(10)

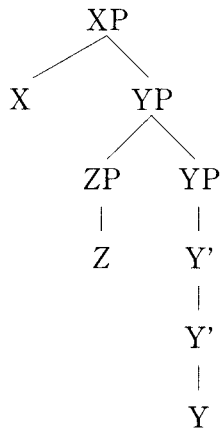


Let X, Y, Z dominate terminals x, y, and z, respectively. Then T={x, y, z} and A={⟨X, Z⟩, ⟨X, Y'⟩, ⟨X, Y⟩, ⟨ZP, YP⟩, ⟨ZP, Y'⟩, ⟨ZP, Y⟩}. d(A)={⟨x, z⟩, ⟨x, y⟩, ⟨z, y⟩}=a total ordering of T.

3. To simplify the proof, I ignore Z' inside ZP as it has no effect on the well-formedness of the structure. Nor do I consider the possibility of Y taking a phrasal complement WP, again because the conclusion would be identical whether the complement exists or not. The reader can prove both omitted cases him/herself.

Second, there is in fact no limit to how many Y' nodes are permitted:

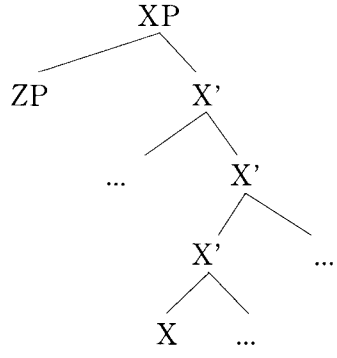
(11)



If all Y' nodes are segments of the same full node, then only the full Y' node (i.e. the set of all Y's) enter c-command, and (11) is practically the same as (10), which has a full Y' node containing a single segment. But even if all Y's are full nodes by themselves, d(A) for (11) is still the same as the one for (10), because for each extra Y'_i different from all other Y's, we only need to add the pairs $\langle X, Y'_i \rangle$ and $\langle ZP, Y'_i \rangle$ to A. Since all Y's dominate the same terminal(s), they have no effect on d(A), the union of what all such pairs in A dominate.

With these two properties, the Y' nodes in (11) strongly resemble the intermediate nodes in the standard X'-template in (12):

(12)



Putting aside the topmost X' (i.e. the sister of the ZP specifier) for now, the intermediate X' nodes both co-exist with ZP and are indefinite in number. The only difference between (11) and (12) is, of course, that X's in the latter are typically created to host adjuncts (in the "..." positions) whereas Y's in the former reject any sister according to the LCA. But because they are so similar otherwise, and because the very existence of the Y's in (11) is pretty mysterious, it seems

attempting to re-consider the status of Y's in Kayne's theory.

Recall that for Kayne, Y' cannot have a phrasal sister (e.g., an AdvP) because otherwise A would contain both ⟨Y', Adv⟩ and ⟨Adv, Y⟩, which in turn would put both ⟨y, adv⟩ and ⟨adv, y⟩ in d(A) and thus make d(A) no longer asymmetric (cf. (6) above). This reasoning is based on the assumption that Y', as a full node, enters into asymmetric c-command with other full nodes when A is computed. Kayne (1994) does not explain why this must be the case. But judging from the context, his consideration seems to be as follows: the LCA is in nature a mathematical (hence abstract) condition on mapping from one set of ordered pairs of nodes (nonterminals under asymmetric c-command) to another (terminals under linear precedence). Thus, if the LCA refers to nonterminal nodes, then ALL nonterminals must be treated equally. Put differently, he tries not to add the ad hoc condition that certain phrasal nodes enter into (asymmetric) c-command while certain others don't. It follows that Y' must be included in the computation of A in the same way that any other phrasal non-terminals are.

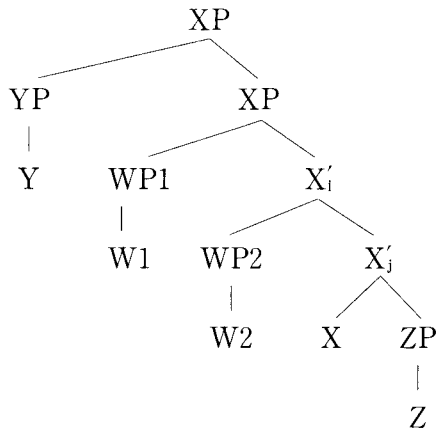
While this is definitely the most general position one can take, it is not maintained even in Kayne's own theory. Remember that, to permit the Specifier in a phrase, Kayne proposes to replace the intermediate node with a segment of the maximal projection (cf. (7)), together with the assumption that c-command does not include segments. But from the mathematical point of view, a node is a node is a node, whether it is a segment or a full category. So in its abstract sense, the LCA should not be sensitive to the full category-segment distinction either. In other words, in his actual implementation of the LCA, Kayne already brings linguistically relevant information into his decision on what non-terminals are selected for c-command. But then there is really no conceptual reason to exclude segments from c-command but at the same time to necessarily include the intermediate Y'.

In an attempt to incorporate the essence of Kayne's LCA into his "bare phrase structure", Chomsky (1995) proposes that of all nontermimals, only maximal projections and heads are "visible" to the computation of (asymmetrically) c-commanding pairs. This is doubtless a linguistically motivated restriction, based on the fact that in the current model of syntax, there is no clear evidence that intermediate nodes ever function as syntactically "useful" c-commanders or c-commandees. But as we saw above, such a restriction on non-terminals is of the same nature as Kayne's one on segments. So let's add Chomsky's proposal to Kayne's definition of c-command in (1), yielding (13):

(13)X c-commands Y iff X and Y are either heads or maximal projections, X and Y are full nodes, X excludes Y, and every full node dominating X dominates Y.

Now consider the structure below, again ignoring the structural details of the specifier, complement, adjunct phrases. All X' nodes are full nodes:

(14)



To illustrate how (13) works, first consider WP1 and WP2. By definition, X'_i doesn't c-command, eliminating the possibility of creating a d(A) that is not asymmetric (cf. (6)). The same X'_i also guarantees that WP1 asymmetrically c-commands WP2 (and hence W2), because X'_i is still a full node that dominates WP2 but not WP1. Therefore, <WP1, W2> is in A but <WP2, W1> is not, which in turn includes <w1, w2> as the only pair involving w1 and w2 in d(A). I leave it to the reader to prove that for (14), this result ultimately leads to a d(A) that is a total order on terminals.[4]

To summarize so far, once intermediate nodes are excluded from the domain of c-commanders and c-commandees, (14) is well-formed under the LCA. This modification does not add any conceptual burden to Kayne's original theory, but it does avoid the mysterious property of it -- that intermediate nodes are permitted only when they have no sisters. (14) also maintains a structural distinc-

4. It is worth noting that the X' nodes may seem to have a double nature. On the one hand, they don't enter c-command because they can't serve as c-commanders or c-commandees. On the other, they do enter c-command because their existence determines the c-command relations between nodes dominated by them. But there is no conceptual inconsistency here. X' nodes are relevant to c-command because they are part of the structural hierarchy that underlies c-command. But being part of the hierarchy does not necessarily turn X' nodes into c-commanders and/or c-commandees.

tion between the specifier position (a typical A-position) and the adjunct positions (A'-positions), a linguistically significant difference that Kayne's theory is forced to blur. Furthermore, the spec position remains unique, as the standard X'-theory dictates, while the adjunct positions are indefinite, another linguistically necessary property of syntactic structures. In fact, (14) is both functionally and hierarchically identical to the standard X'-template in (12), except for two things.

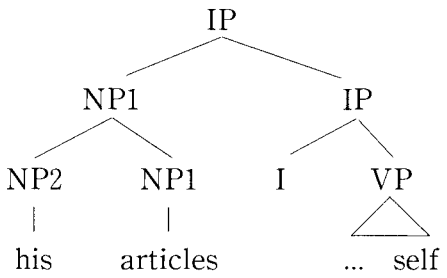
First, because Kayne proposes to order terminals with linear precedence, even adjuncts must linearly precede the heads in (14), whereas the standard (12) has no such restriction encoded in it. The consequences of Kayne's proposal still need substantial investigation. But strictly speaking, the LCA itself does not impose linear precedence on terminals. Even if WP2 is placed linearly after X'_j, the resulting d(A) will still be a mathematically total order, as the reader can prove him/herself. The same point is independently noted by S. Epstein (cf. Chomsky (1995)). For this reason, I will assume the standard assumption about the X'-theory: The template only determines structural hierarchy, with the linear order determined by separate parameters.

The second difference between (14) and (12) is that the specifer is adjoined to the maximal projection in (14) but is a sister to X' in (12). Some desirable consequences of an XP-adjoined specifer are briefly explored in Kayne (1994). From the point of view of Chinese syntax, it is worth mentioning that (14) can readily handle the basic cases that motivate Huang and Tang's (1991) sub-command:

(15)tade wenzhang jingchang chuipeng ziji.
his articles often boast self
Lit. 'his articles often boast about himself'

In order for *tade* 'his' to bind the anaphor ziji 'self', Huang and Tang proposed that binding can be established as long as the binder is immediately dominated by a phrase that c-commands the anaphor. But under (13-14), no such new notion is needed for (15), whose relevant structure is given in (16):

(16)



As the specifier of the subject NP1, NP2 asymmetrically c-commands VP: Both are phrasal full nodes; neither includes the other; every full node dominating NP2 dominates VP because no full node in (16) dominates NP2; and finally, there is one full node, IP, that dominates VP but not NP2. It follows that the binding relation between his and self can be established directly through c-command, dispensing with the need for sub-command, as Kayne noted himself (p. 136).

Furthermore, (13) predicts, correctly, that a more deeply embedded antecedent will fail to license the anaphor:

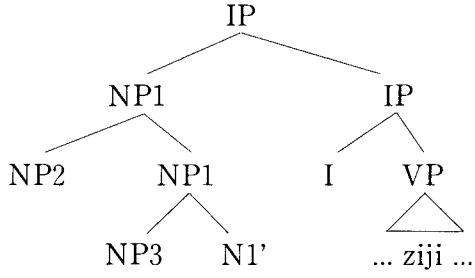
(17) * [[PP youguan Baoyuk] de xiaoxi] shi zijik-de jiaren hen shangxin.
about Baoyu 's info make self's family very sad
'The information about Baoyu made his family members sad'

(18) * [[CP Baoyuk chujia] de xiaoxi] shi zijik-de jiaren hen shangxin.
Baoyu become-a-monk 's news make self's family very sad
'The news that Baoyu became a monk made his family members sad'

Since *Baoyu* is dominated by a full PP node in (17) and at least by a full CP node in (18), it doesn't c-command the anaphor *ziji* under (13), making the examples unacceptable.

The theory proposed in this short article also predicts a more subtle contrast between specifiers and adjuncts in the role of anaphor binding, a contrast unexpected from Kayne's original set of definitions in (1)-(3). Consider the abstract structure below:

(19)



NP2 is the specifier of the subject NP1, and NP3 is the adjunct. As shown above, (13) permits NP2 to c-command the anaphor ziji in VP. In contrast, NP3 is dominated by a full node, NP1, which doesn't dominate ziji, so no binding relation can be established. The data is subtle, but seems to conform to the prediction:

(20)Baoyu de xiaoxi shi Wang Furen hen shangxin.
Baoyu'd info make Wang Lady very sad
'The info Baoyu gave made Lady Wang sad' or
'The info about Baoyu made Lady Wang sad'
(21)Baoyu$_k$ de xiaoxi shi ziji$_k$ de jiaren hen shangxin
Baoyu 's info make self 's family very sad
'The info Baoyu gave made his family members sad'
??/*'The info about Baoyu made his family members sad'

(20) shows that *Baoyu de xiaoxi* 'Baoyu's info' is potentially ambiguous, depending on whether the NP *Baoyu* is treated as the possessor in the Spec position or as an adjunct comparable to the PP in (17).[5] When *Baoyu* is forced to serve as the binder for an anaphor in (21), the ambiguity seems to disappear, with the adjunct reading extremely difficult (or impossible for some people) to obtain. This specifier-adjunct distinction is not predicted by Kayne's theory, which assigns the same position to the specifier or adjunct because that is the only XP-internal phrasal position different from the complement. (Again, the reader can prove this by her/himself.) It follows that if the specifier NP can c-command the anaphor, so can the adjunct NP, contrary to the Chinese data in (20-21).

To conclude, the reformulation of c-command in (13) both removes a theoretical oddity in Kayne's theory of the LCA and has a firmer empirical basis. Meanwhile, it retains the original advantage of subsuming sub-command.

Appendix: More on Nodes Qualified as C-commanders

Taking up my observation that the nonterminals entering c-command should be determined linguistically instead of mathematically as in Kayne's original attempt, a reviewer wonders if c-command can be further modified as follows:

(22)X c-commands Y iff X is a maximal projection, X and Y are full nodes, X excludes Y, and every full node dominating X dominates Y.

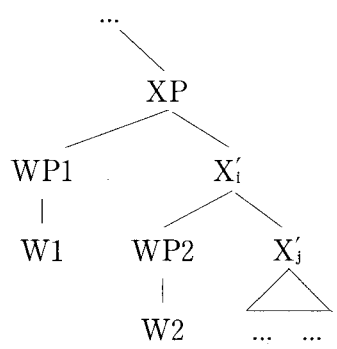
That is, only phrasal nodes are c-commanders whereas all full nodes, regardless of their projectional level, can be c-commandees. In this appendix, I will first try

5. Evidence that no prenominal phrase in an NP is in the complement position is presented in Li (1997a).

to explicitly spell out this reviewer's suggestion. Then I will discuss a complication that prevents it from being fully developed in the current paper.

First, (22) has no empirical difference from (13) with respect to X'. This can be shown with a subtree of (14):

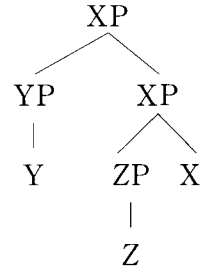
(14)



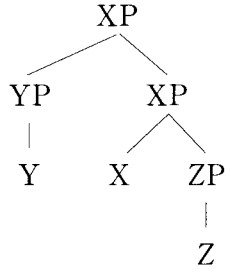
In order to permit X' nodes in the LCA model, it is only crucial to prevent X' from c-commanding its sister so that, for intance, $\langle X'_i, W1\rangle \notin A$. Whether X'_i can be c-commanded by WP1 is inconsequential provided that WP1 asymmetrically c-commands WP2 and X'_j, a condition that is easily met as long as X's count as full nodes. Therefore, (22) permits adjuncts for the same reason that (13) does, as the reviewer correctly suspected.

Now consider heads as non-c-commanders. The reviewer observes that this will make the head-final word order a default in human languages if d(A) indeed represents the temporal precedence relation as in Kayne's original proposal (cf. (1)-(3) above). This, again, is correct. Compare (23a) and (23b), with x, y, z being terminals for X, Y, Z, respectively:

(23) a.



b.



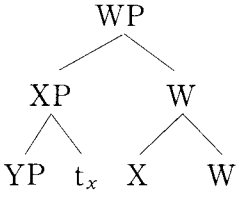
In both cases, the specifier YP asymmetrically c-commands the complement ZP as before. It follows that $\langle YP, Z\rangle \in A$ for each tree, which in turn makes $\langle y, z\rangle \in$ d(A). This means the specifier temporally precedes the complement under Kayne's assumption summarized in (2b) above. Furthermore, since heads$\neq$c-commanders under (22), $\langle ZP, X\rangle$ is in A but $\langle X, ZP\rangle$ isn't, making $\langle z, x\rangle \in$d(A)

and $\langle x, z\rangle \in d(A)$. This means only (23a) is compatible with the LCA because the terminal x temporally precedes the terminal z in (23a) but follows z in (23b). As a result, we conclude that the head-final structure is the only one permitted by the LCA under (22), once temporal precedence is taken into consideration. As the reviewer notes, this result is "convergent with that of Takano (1996)", who uses a reverse operation of Merge at PF to achieve the same result.

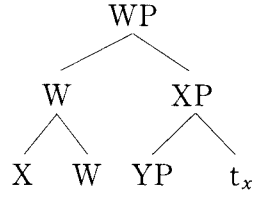
It should be pointed that the idea of excluding heads as c-commanders cannot be simply dismissed because c-commanding heads seem to play such indispensable roles in the principles-and-parameters model of syntax. As an example, since the ECP relies on some strong version of government, typically defined in terms of c-command by a lexical head, it may seem impossible to prevent heads from c-commanding. However, Li (1997b) argues that the head-government half of the ECP is actually a misled conception, that it can and should be reduced to the more fundamental X'-licensing conditions between heads and their complements/specifiers. The specific solution in that article cannot be directly incorporated into the LCA-model, but it does show that it is highly suspicious to regard c-command by a lexical head as the right condition for trace-licensing, thereby conceptually removing one of the few strongest obstacles to (22).

The most obvious difficulty with adopting (22) lies in permitting the SVO word order. As we saw above, this definition of c-command only generates the SOV sequence once d(A)=the relation of temporal precedence on the set of terminals. In Takano (1995), the SVO sequence results from verb-raising. But this can't be done in any simple manner with (22). Consider the two alternative structures below (irrelevant details omitted):

(24) a.



b.



(24a) consists of consistently head-final phrases, while the upper phrase in (24b) is head-initial. First note that as long as the raised head X, adjoined to W, is a head, (24b) is ruled out by the LCA because XP asymmetrically c-commands X. Since XP dominates YP and $\langle XP, X\rangle \in A$ $\langle y, x\rangle \in d(A)$. And this contradicts the precedence relation encoded in (24b), where x temporally precedes y. In contrast, (24a) is well-formed by the LCA. Unfortunately, the resulting linear order remains head-final.

As part of the minimalist program, Chomsky (1995, 1996) proposes that a

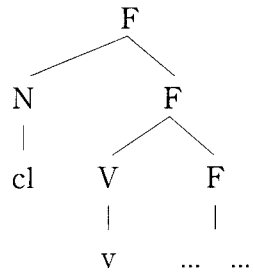
head be regarded as a simultaneous maximal projection if it no longer projects. For the purpose of the current discussion, let us temporarily assume that such a head must be regarded as a maximal projection.[6] It follows that X in (24) is a maximal projection which can indeed c-command according to (22), as the reviewer pointed out. This makes X c-command XP and asymmetrically c-command YP/Y since only a segment of W separates X from XP. Therefore, $\langle X, Y\rangle \in A$ and $\langle x, y\rangle \in d(A)$. Considered alone, this is a desirable result because the LCA now requires that X temporally precede XP, the complement of W. And the only way to accomplish this is to abandon (24a) and adopt (24b). In other words, raising of X to a higher head position indeed generates the VO sequence, with X=V and YP=the object NP. However, (24b) violates the LCA for a different reason: With W unable to c-command, $\langle XP, W\rangle \in A$. Since XP dominates YP, this ordered pair of non-terminals makes $\langle y, w\rangle \in d(A)$. But this is impossible because w precedes y in (24b), not the other way around.

This last problem can be avoided if the terminal item under W is always a phonetically null morpheme. A soundless item is arguably not subject to the LCA because temporal precedence can never be "realized" between such an item and other terminals. This solution works naturally when W is a light verb without any phonetic content, as Chomsky (1996) suggested in the context of a Larsonian shell. It also works if W is a functional head originally filled with nothing but a few features. Suppose that X is V and already carries the Tense/Agr morphology when it enters syntactic derivation (cf. Chomsky (1996)). Then raising X to W (via the light verb) will force a structure like (24b), in which X satisfies the LCA because it is a maximal projection and W is permitted because it dominates a soundless terminal item that is not subject to the LCA.

Romance cliticization has often been analyzed as adjoining a nominal head to some functional head. The result follows with X being the clitic and W being the empty functional head it adjoins to. The complication lies in whether another lexical terminal, a verb for instance, also adjoins to the same functional head. If it does, the LCA prevents the following "conventional" configuration, as Kayne already pointed out given his own set of definitions.

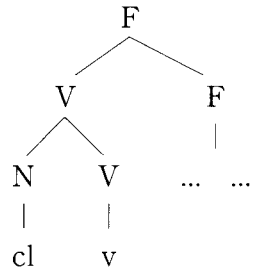
6. Under this assumption, no head can vacuously project to a separate phrasal node as Kayne does with, say, the ZP in (6). See Chomsky (1995) for a way to convert such phrases to the minimalist "bare" phrase structures.

(25)



Because both N and V count as maximal projections and because they are separated from each other only by segments of F (= the functional head), neither ⟨N, V⟩ nor ⟨V, N⟩ is in A, so neither ⟨cl, v⟩ nor ⟨v, cl⟩ is in d(A), making the latter not a total order of T, the set of all terminals in the tree. Kayne explored one alternative:

(26)



As we saw in section 2, N c-commands V because it excludes V, but V doesn't c-command V for lack of exclusion. Therefore, ⟨N, V⟩∈A and ⟨cl, v⟩∈d(A). Furthermore, since N and V c-command every non-terminal X that is dominated by the immediate projection of F, they must be placed before X, generating the head-initial sequence. But there is another theoretical possibility: The clitic and the verb never move to the same functional head. When the clitic raises, it adjoins to a unique functional head. This will require that we have more functional projections in a clause than normally assumed. Making a choice between these two alternatives is clearly beyond what this article can do, but no fundamental problem arises, as far as I can see.

One type of construction that may seem to resist this analysis is incorporation, as discussed in Baker (1988) and many subsequent works. Incorporation must be considered because many languages with this phenomenon do demonstrate the SVO word order. If Baker is correct, then W in (24) must be a phonetically non-empty morpheme (e.g. a causative suffix), thereby posing a problem for my solution. However, it is possible that incorporation is not formed

in syntax in the way Baker suggested.[7] Suppose that all morphologically complex words are indeed formed in the lexicon, as the lexicalist model has suggested. But because they are morphologically complex, they contain more features than morphologically simple words. In order to check off these "extra" features, the syntactic structure in which such a word occurs must have more head and/or Spec positions, as the minimalist program dictates. Thus, a V-caus complex needs a bi-clausal structure to check off, say, the verbal features that each verbal morpheme carries, etc. But because the V-caus complex is the only phonetically non-empty terminal in V-incorporation, the matrix V position it raises to is occupied only by a light verb. The LCA works as before, while the typical bi-clausal properties in the V-incorporation construction also follow naturally.

This treatment of incorporation is in fact supported by two phenomena investigated in Li (1990), namely, that verb incorporation is triggered only by causative verbs and modal-like verbs, and that no functional morphemes are part of the incorporated words even though they are expected in Baker's theory, in which the embedded verb stem raises to the matrix verb via embedded I and C (also see Baker (1996) for further confirming these observations). On the basis of these facts, I proposed the generalization that verb incorporation takes place only when the matrix verb selects a VP complement rather than a CP complement, and explained this generalization along the line of improper movement. But there is an alternative (and simpler) explanation for the facts. It seems most natural that there is direct correspondence in syntactic and morphological selections -- If X selects YP as complement in syntax, then X may select Y as complement in morphological processes. It follows that CP-selecting verbs can never select a verb as complement in morphology. Nor can they select the functional head they actually expect. In the current syntactic theory, functional morphemes don't attach "up" to the matrix verb; they attach "down" to the verb that would eventually be in the same clause. So only VP-taking verbs can successfully select a verbal stem as complement. Furthermore, the cluster thus formed in the lexicon can only take one set of functional morphemes, those to be checked off in the matrix clause. If they take two sets, one for the matrix clause and one for the complement, then the complement set can't be checked off because there isn't

7. A strongly lexicalist treatment of incorporation can be found in Di Sciullo and Williams (1987). But they fail to account for many facts in nearly as elegant ways as Baker does. My suggestion here may be regarded as a sketchy attempt to combine these two approaches. There are other possible solutions. E.g., once X and W forms the inseparable cluster X-W, it may be possible that (24b) satisfies the LCA if one of the components of the cluster does.

any functional head in the syntactic complement phrase to do the required feature checking, with the complement being a bare VP by definition. While this analysis follows from the most basic assumptions of the current model of syntax, it crucially relies on formulating incorporation in the lexicon, and it works well with my proposal on LCA-based head-movement, under (22).

Now a few questions to be answered. First, my proposal depends on treating raised heads as maximal projections. This assumption not only allows the raised head to satisfy the LCA in the desirable linear order, but also allows it to be a c-commanding antecedent for the trace. But this last virtue is also a potential problem: Can a head trace form a chain with a phrasal antecedent? In Chomsky's particular formulation of the minimalist program, such a chain is illegitimate at the interface level. But it is not clear that Chomsky's formulation is sufficiently motivated. Clearly, more evidence is needed to make a choice.

Second, Pollock (1989) argues that French V moves to Agr but English V doesn't. More generally, we certainly don't have evidence that every SVO language involves moving V all the way to C. Then how do we account for the Spec-head-complement sequence in these languages when the head is functional. The situation may not be as insurmountable to my proposal as it appears. Even if V doesn't move to a functional head, Agr and Tense always seem to occur together. If they involves head-raising, then the resulting structure becomes head-initial like (24b).[8] As for complementizers, we may have to reconsider their true status: Maybe the English *that* actually occupies the SpecCP position, particularly considering that there is no satisfactory explanation for the doubly-filled-COMP filter in the principles-and-parameters model. Again, the issue is too complex to be dealt with in this short article, though various solutions may be explored.

Lastly, the linear locations of phrasal adjuncts will need more careful examination. By default, they should always precede the complement in the strict head-final structure. Because this issue is closely tied to how far lexical heads and complement phrases move in a given language, it simply cannot be dealt with here.

Because of these complications, I have decided not to incorporate (22) into the more modest modification of Kayne's theory of phrase structures which I laid out in section 2. But I hope to have shown that it is at least worth serious investigation, probably along the lines I proposed in this appendix.

8. See Kayne (1994) for treating functional suffixes in SOV languages as phonological clusters instead of syntactic ones.

References

Baker, Mark. 1988. *Incorporation: A Theory of Grammatical Function Changing*. Chicago: Chicago University Press.

Baker, Mark. 1996. *The Polysynthesis Parameter*. New York: Oxford University Press.

Chomsky, Noam. 1986. *Barriers*. Cambridge, MA: MIT Press.

Chomsky, Noam. 1995. "Bare Phrase Structure." In Gert Webelhuth, (ed.) *Government and Binding Theory and the Minimalist Program. Cambridge, MA: Blackwell.*

Chomsky, Noam. 1996. The Minimalist Program. Cambridge, MA: MIT Press.

Di Sciullo, Anne-Marie and Edwin Williams. 1987. *On the Definition of Word*. Cambridge, MA: MIT Press.

Huang, James and Jane Tang. 1991. "The Local Nature of the Long-Distance Reflexive in Chinese". In Jan Koster and Eric Reuland, (eds.) *Long-Distance Anaphora*. p. 263-282. Cambridge: Cambridge University Press.

Kayne, Richard. 1994. *The Antisymmetry of Syntax*. Cambridge, MA: MIT Press.

Li, Yafei. 1990. "Verb Incorporation and X°-binding." *Linguistic Inquiry* 21:399-426.

Li, Yafei. (1997a). "Remarkss on Chinese Word Order." *Journal of Chinese Linguistics* 1.1.

Li, Yafei. (1997b). "Head-government and X'-theory." *The Linguistic Review* 14.1.

Pollock, Jean-Yve. 1989. "Verb Movement, Universal Grammar, and the Structure of IP." *Linguistic Inquiry* 20:365-424.

Takano, Yuji. 1995. *Movement and Parametric Variation in Syntax*. Ph.D. dissertation, UC-Irvine.

（編輯部按：本文實際出版日期爲 1998 年 3 月。）

線性相關定理與標準 X'-理論的關係

李亞非
威斯康森州立大學麥迪森分校語言學系

摘　要

Kayne 于 1994 年提出的線性相關定理，給人的印象是在句子結構裡不准有次詞組節點 X'。事實上，他的理論仍舊允許 X'，但是這個節點存在的必要條件是它在句子結構裡不起任何劃分層次的作用。該結果與節點的一般性特徵不符。解決問題的方法是把 X'節點排除在 c-command 關係之外。修改後的理論既保持了原有理論的全部主要優點，也從結構上避免了原有理論不能區分指定成分和修飾成分的缺點。X°- 移位所遇到的一些問題，也有可能沿著這個理論的基本思路找到答案。

關鍵詞：線性相關定理，X'-理論，X°-移位