On the Connection between Hierarchy and Order

Marcus Kracht

II. Mathematisches Institut
Freie Universität Berlin
Arnimallee 3
D-14195 Berlin
`kracht@math.fu-berlin.de`

February 8, 2008

1 Introduction

Recent developments in syntactical research have shown a great convergence in the view that hierarchy and order are not independently specifiable and that preferrably one is determined by the other. Furthermore, in connection with this there is growing consensus that there must be a fundamental word-order in (the nowadays obsolete) D-structure from which the surface word order is derived by successive movements. The consensus ends there, however. For there are at the moment two views on the matter, which are in direct conflict. On the one hand there is [2] who argues that the word order ov is fundamental, on the other hand there is [**?**] who in effect derives from his theory that all languages are fundamentally vo. We will leave it to empirical research to give evidence which of the theories is to be preferred or whether both must be dismissed. Here we will concentrate on the purely mathematical evaluation of these theories. This might prima facie be not such an interesting thing to do. [**?**] has anyway indicated what consequences his theory has and in addition [2] is as clear in this respect as can be expected. However, let us not forget that every proposal will sooner or later be modified to enlarge the descriptive potential or remedy certain shortcomings. To give just one example, [**?**] has modified the definition of c-command in order to allow for two adjunct-positions rather than one in Kayne's theory. Obviously, this will not be the last modification and the question arises as to what theries of sentence structure can be associated with a linear correspondence principle à la Kayne.

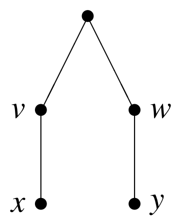
Secondly, we must ask ourselves what exactly entitles us to deduce that there is such a thing as a strict basic word order. This is not so easily seen as the current discussion makes us believe. It can be observed that a restriction of one component of grammar typically goes hand in hand with opening up another. For example, the theories of word order that we just mentioned concern themselves with the structure inside a maximal projection. At the moment the consensus is lost as to how many basic categories there are and what their fundamental ordering in languages is. Obviously, the differences in word order that we do observe must be accounted for. If a thing such as the directionality parameter is abolished, then either we assume different behaviour of elements under movement (generally assumed not to be a good choice) or we assume that the languages select different fundamental alignments of basic categories. To give an example, [3] proposes that English generates the NEG-phrase below the TNS-phrase whereas Basque generates the TNS-phrase below the NEG-phrase. Modern conceptions in the minimalist framework solve the problem of surface aligment by proposing a difference in strength of morphological features. A priori these seem to be just different ways of attacking the same thing. Apart from other problems that the different approaches solve *in addition*, there is nothing that singles out one of them as the most ideal solution. Maybe, then, it is a good idea to study abstractly the potential that lies in each of these approaches so that (i) one may have theoretical reasons for preferring one over the other if they turn out to be different or (ii) one may express any theory in any of these frameworks if they are equal and choose whichever suits best the current problem. For a linguist, (ii) seems a less ideal situation because it fails to show that one of the approaches is better. But if they are in fact the same this will show that there is no way to choose between them empirically and one should give up trying to do so. Admittedly, we will never reach a point where such results can be proved because a theory cannot really be robbed of its connotations and one will always feel that one is more adequate than the other despite the fact that they are pravtically identical.

2 An Outline of the Theory of Kayne

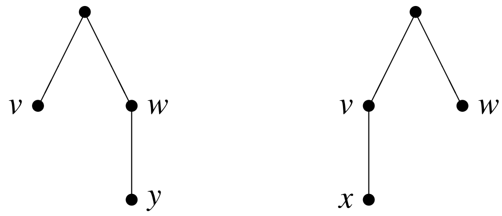
Kayne assumes that there is a direct correspondence between the linear order of the words and the asymmetrical c-command relation in trees; this correspondence is expressed by the *linear correspondence axiom* abbreviated as *LCA*. We will explain this axiom in its most simple version on trees without the segment/category distinction. Recall that in a tree a node x *c-commands* a node y if the first node immediately above x dominates y but neither does x dominate y nor does y dominate x. x *asymmetrically c-commands* y if x c-commands y but y does not c-command x. Now let x, y be leaves (= terminal nodes) of the tree. We write $x \sqsubset y$ if there exist non-terminal nodes x', y' such that $x' > x$, $y' > y$ and x' asymmetrically c-commands y'.

LCA. $\sqsubset$ is a strict linear order.

Moreover, Kayne argues that $\sqsubset$ is not just any linear order but in fact *precedence in time*. In the simplified example it can be shown that the LCA reduces to the requirement that structures must be strictly right branching. Consider, namely, what happens if the following situation arises.



Here, v c-commands y but y does not c-command v, so v asymmetrically c-commands y. Likewise, w asymmetrically c-commands x. Let now x dominate a terminal node p and y a terminal node q. Then we have both $p \sqsubset q$ and $q \sqsubset p$ but $p \neq q$, so LCA is violated. This leaves us with only the following possibilities for subtrees.



Moreover, if we assume that $\sqsubset$ is temporal precedence and temporal precedence is the left-to-right order on paper then the second of the trees is excluded as well. Consequently, the structures satisfying LCA are strictly right branching trees.

Now Kayne's theory isn't as simple minded as that. Rather than using the c-command relation specified above it uses the adjunction-based definition given in [1]. In order to fully appreciate this definition and its empirical content it is necessary to spell out exactly what *adjunction structures* are.

3 Trees and Adjunction Structures

While trees are fundamental structures both in mathematics and linguistics they are rigorously defined and there is no problem in handling definitions which are strictly tree based. However, ever since the barriers system, nodes in trees do no longer function as carriers of categories. This shift in the fundamental notions has caused a great deal of confusion since it was not properly mirrored by a shift in terminology. Moreover, the notions are at closed look not as rigorously defined as they should be. In this section we will try to remedy this situation.

Recall that a (finite) **tree** is a structure $\mathfrak{T} = \langle T, r, < \rangle$ where T is a (finite) set, the set of **nodes**, $r \in T$ a node, called the **root** and $<\subset T \times T$ a binary relation such that the following conditions hold

Transitivity	$(\forall xyz)(x < y \wedge y < z. \to .x < z)$
Irreflexivity	$(\forall x)\neg\ x < x$
Local Linearity	$(\forall xyz)(x < y \wedge x < z. \to .y < z \vee y = z \vee y > z)$
Root	$(\forall x)(x = r \vee x < r)$

Given a tree, we define $x \leq y$ by $x = y$ or $x < y$. If $x \leq y$ we say that y **dominates** x, if $x < y$ we say that y **properly dominates** x. Finally, if there is no z such that $x < z < z$ we say that y **immediately dominates** x. Given x we put

$$\begin{array}{rcl} \downarrow x & = & \{y : x \geq y\} \\ \uparrow x & = & \{y : x \leq y\} \end{array}$$

$\downarrow x$ is the **constituent** headed by x and $\uparrow x$ the **position** of x. Next we discuss *ordered trees*. An **ordered tree** is an object $\mathfrak{T} = \langle T, r, <, \sqsubset \rangle$ such that $\langle T, r, < \rangle$ is a tree and $\sqsubset \subset T \times T$ a binary relation on the nodes equivalent to the order *in time*. This relation is not so straightforward to define since we have to take of overlapping nodes. The intuition is that $x \sqsubseteq y$ for two nodes x, y if the event of uttering x precedes that of y. It is clearly possible for events x, y that neither does x precede y nor does y precede x; in that case we say that they *overlap* and write $x \circ y$. This results in the following postulates.

Definition	$(\forall xy)(x \circ y. \leftrightarrow .\neg\, x \sqsubset y \wedge \neg\, y \sqsubset x)$
Transitivity	$(\forall xyz)(x \subset y \wedge y \sqsubset z. \rightarrow .x \sqsubset z)$
Irreflexivity	$(\forall x)\neg\, x \sqsubset x$
Overlap	$(\forall xy)(x \circ y. \leftrightarrow .x \geq y \vee x \leq y)$
Persistence	$(\forall xyx'y')(x \sqsubset y \wedge x' \leq x \wedge y' \sqsubset y. \rightarrow .x' \sqsubset y')$

Given the interpretation these postulates are immediate. If x precedes y and y precedes z then x precedes z as well; x cannot precede itself. If x does not precede y nor does y precede x then they must overlap. Here, however, we have the special situation of trees. Overlapping events must be included in each other, that is, if x overlaps with y then either x is included in y or y is included in x. This inclusion (as subevent) manifests itself as dominance in the tree. The last postulate concerns the consistency of the precedence relation with respect to subevents. If x precedes y and x' is a subevent of x, y' a subevent of y then x' must precede y' as well.

Finally, we come to *adjunction structures*. The idea behind adjunction structures is that they arise from a process known as *adjunction*. The effect of adjoining a constituent to a node x is that x becomes doubled up into two nodes; so linguistic categories do not correspond to nodes but to sets of nodes which arise from adjunction. It is not hard to see that if the adjunction to x creates the new node *above* x then the sets corresponding to categories must be linear. Thus we have the following definition.

Definition 1 *An* ***adjunction structure*** *is an object* $\langle T, r, <, C \rangle$ *where* $\langle T, r, < \rangle$ *is a* *tree and $\mathbb{C} \subseteq Pow(T)$ a partition of the set T of nodes into convex, linear subsets. An element $\mathfrak{b}$ of C is called a* ***block****, and $x \in \mathfrak{b}$ is a* ***segment*** *of* $\mathfrak{b}$.

Similarly ordered adjunction structures can be defined. Recall that a *partition* of a set T is a set $\mathbb{C}$ of subsets such that if $\mathfrak{b}_1, \mathfrak{b}_2 \in \mathbb{C}$ then $\mathfrak{b}_1 \cap \mathfrak{b}_2 = \emptyset$ and the union of all $\mathfrak{b} \in \mathbb{C}$ is the full set T. Thus each node is a segment of exactly one block. A set S is called *convex* if for all $x \leq y \leq z$ $y \in S$ if only $x, z \in S$. S is *linear* if for two elements x, y either $x = y$ or $x < y$ or $x > y$. A block is always of the form $[x, z] = \{y : x \leq y \leq z\}$.

The notions of dominance etc. have to be adapted to adjunction structures. There are now three possibilities for blocks to dominate each other. Take namely two blocks, $\mathfrak{b}_1$ and $\mathfrak{b}_2$. Let $x \in \mathfrak{b}_1$ and $y \in \mathfrak{b}_2$. If $x > y$ then for all $z \in \mathfrak{b}_2$ $x > z$. For by the linearity of $\mathfrak{b}_2$ we must have $z = y$, $z < y$ or $z > y$. Only the last case is nontrivial. We have $x > y$ and $z > y$; thus by *local linearity* either $x > z$ or $x = z$ or $x < z$. In the first case we are done. The last two, however, cannot arise. For if $x \leq z$ and $y, z \in \mathfrak{b}_2$, then by the convexity of $\mathfrak{b}_2$ we have $z \in \mathfrak{b}_2$ and so $x \notin \mathfrak{b}_1$, since $\mathfrak{b}_1$ and $\mathfrak{b}_2$ are disjoint.

Given $\mathfrak{b}_2$, the set of nodes in $\mathfrak{b}_1$ which dominate one (and therefore all) segments of $\mathfrak{b}_2$ is linear and convex. Let namely $\mathfrak{b}_1 = [x, y]$ and let z be the least segment dominating a segment of $\mathfrak{b}_2$. Then the nodes of $[z, y]$ will be all nodes from $\mathfrak{b}_2$ dominating a segment of $\mathfrak{b}_2$. There are three cases. (1) z does not exist, (2) $z < y$, (3) $z = y$. If (1) is the case we say that $\mathfrak{b}_1$ **excludes** $\mathfrak{b}_2$; if (3) is the case we say that $\mathfrak{b}_1$ **includes** $\mathfrak{b}_2$. If (2) is the case we say that $\mathfrak{b}_1$ **weakly includes** $\mathfrak{b}_2$.

There is the possibility to reduce a category to a single node; thus we realize the intuition that categories are represented by a single node. In this case we need to use the relations of *inclusion* and *weak inclusion* as primitive. Let us then define an **adjunction tree** as an object $\mathfrak{A} = \langle A, r, \ll, \leq \rangle$ where the nodes of $\mathfrak{A}$ are the blocks of an adjunction structure and $a \ll b$ if a is included in b and $a \leq b$ if a is wekaly included in b. First of all we should ask ourselves whether we can characterize adjunction trees independently. This is possible; the postulates are the following.

Irreflexivity	$(\forall x)(\neg\, x \ll x)$
	$(\forall x)(\neg\, x \leq x)$
Transitivity	$(\forall xyz)([x \ll y \vee x \leq y] \wedge y \ll z. \rightarrow .x \ll z)$
Persistence	$(\forall xyz)([x \ll y \vee x \leq y] \wedge y \leq z. \rightarrow .x \leq z)$

First question is whether these postulates indeed determine adjunction structures, that is, whether any structure satisfying these postulates can be construed as the compression of an adjunction structure. To this end take any adjunction tree $\mathfrak{A} = \langle A, r, \ll, \leq \rangle$. Crucial is here the relation $(\ll \cup \leq)^+$; this relation plays the part of dominance. Start with the root and work your way down. The root is now a block; we need to see how many nodes are to be put into that block. First of all, collect all elements which are *immediately included* by r with respect to the relation $(\ll \cup \leq)^+$. They can be divided into the set $I = \{i_1, \ldots, i_m\}$ of elements included by r and the set $W = \{w_1, \ldots, w_n\}$ of elements weakly included (but not properly included). Now replace r by two copies r_w and r_i and let r_w immediately dominate W in any order plus r_i; let r_i immediately dominate I in any order. Now proceed by splitting i_j, w_k in the same way. It is not hard to see—given the success of this uncompression—that the adjunction structures do not let us recover the adjunction structure uniquely. The problem is not the set I but the set W. We cannot single out the correct hierarchical structure concerning the adjoined element, that is, the weakly included elements which are not strongly included. We cannot say of two such elements v, w whether they are sisters or which of the two is higher in the sense of strict c-command. Similarly, if we consider the ordered companions, the recovery is not unique. The additional postulates for the ordering relation (and overlap) are the following in addition to the intrinsic postulates *Irreflexivity*, *Transitivity* and *Overlap*.

R-Persistence	$(\forall xyz)(x \sqsubset z \wedge [y \ll z \vee y \leq z]. \rightarrow .x \sqsubset y)$
L-Persistence	$(\forall xyz)(x \sqsubset z \wedge [y \ll x \vee y \leq x]. \rightarrow .y \sqsubset z)$

We could have formulated a single postulate but two are more readable. Now consider an ordered adjunction structure and two blocks $\mathfrak{b}$ and $\mathfrak{c}$. Then either one segment of $\mathfrak{b}$ dominates or is dominated by a segment of $\mathfrak{c}$ in which case they overlap, or no segment of $\mathfrak{b}$ dominates any segment of $\mathfrak{c}$ and no segment of $\mathfrak{c}$ dominates any segment of $\mathfrak{b}$. In case of overlap, if a segment of $\mathfrak{b}$ dominates a segment of $\mathfrak{c}$ then it dominates *all* segments of $\mathfrak{c}$. Thus overlap means that the event of $\mathfrak{b}$ is included in or includes the event of $\mathfrak{b}$ if by the *event* for $\mathfrak{b}$ we understand the set of all nodes (improperly) dominated by *any* segment of $\mathfrak{b}$.

4 Command Relations and Asymmetric Kernels

A **command relation** is a function s selecting for each tree $\mathfrak{T}$ a relation $s(\mathfrak{T})$ over that tree such that the following conditions hold.

Stability	:	If $U \subseteq T$ is a subtree then $s(\mathfrak{U}) = s(\mathfrak{T}) \cap T \times T$
Shape	:	The set $\{y \in T : x\ s(\mathfrak{T})\ y\}, \neq r$ is a constituent properly containing x
Root	:	The set $\{y \in T : r\ s(\mathfrak{T})\ y\}$ is the whole tree

These definitiopns merit comment. First of all, the notion of a subtree of a tree needs to be clarified. We will understand by a **subtree** $\mathfrak{U}$ of $\mathfrak{T}$ an object $\mathfrak{U} = \langle U, s, < \cap U^2 \rangle$ arising from T by restricting the relation of T to a subset U where U is such that it has a unique root s and is convex. (Recall that *convex* means that we have $z \in U$ if only $x < z < y$ for some $x, y \in U$.) The stability postulate has the following cnsequence. Let us pick two arbitrary nodes in $\mathfrak{T}$, x and y. Then whether or not $x\ s(\mathfrak{T})\ y$ can be decided by looking at the minimal subtree containing both x and y. We can get this subtree by taking the least point z dominating both x and y, called the **crosspoint** and taking $U = [x, z] \cup [y, z]$. Subtree stability is an important concept.

Now we come to the postulate for *shape*. Let us agree on the following terminology. We say x **commands** y with respect to s (or $s(\mathfrak{T})$) if $xs(\mathfrak{T})y$. The set of all nodes commanded by x in $\mathfrak{T}$ is called the **domain** of x in $\mathfrak{T}$. Domains are constituents according to shape; this means that the domain of x is of the form $\downarrow y$ for some y. We write $y = f_s(x)$, ignoring the dependency from the tree $\mathfrak{T}$ which we would normally have to indicate. If s is understood we simply write $y = f(x)$. f_s is the **function associated** with s. Thus, instead of defining command relations as systems of relations over trees we could define them as systems of functions over trees. Notice that an additional consequence of *Shape* is that $f(x) > x$ if x is not the root. *Root* is equivalent to $f(r) = r$.

Command relations can similarly be defined on ordered trees. Yet, there is typically no ordering condition on command relations. This is expressed in the postulate of *Ambidextrousness*.

Ambidextrousness : If $\sqsubset_1$ and $\sqsubset_2$ are two orderings of the same tree $\mathfrak{T}$ then $s(\langle \mathfrak{T}, \sqsubset_1 \rangle) = s(\langle \mathfrak{T}, \sqsubset_2 \rangle)$.

This concludes the definition of command relations over trees. However, there are postulates with which we will have to deal extensively later on. These are

Monotonicity	:	if $x \leq y$ then $f_s(x) \leq f_s(y)$
Tightness	:	if $x < f_s(y)$ then $f_s(x) \leq f_s(y)$

Monotone relations have the reasonable property that domains are never shrinking when one moves up the tree. Tight relations have the property that they have a node-based definition. Such a definition works by selecting a set $P \subseteq T$ of critical nodes and then defining $f_P(x)$ to be the least node $y \in P$ strictly dominating x, and $f_P(r) = r$. Given P we say that x **P-commands** z if z is in the domain of that relation.

Given a relation R on a tree—in our case a command relation – the asymmetric kernel $R^{>}$ is defined as follows. $xR^{>}y$ iff (1) $x\,R\,y$, (2) $\neg\, y\,R\,x$ and (3) $\neg\, x \circ y$. The last clause does not necessarily presuppose an ordering on the nodes; we can define $x \circ y$ by $x \leq y \vee y \leq x$. Central to Kayne's analysis of word-order, and indeed central to syntax based approaches to trees, is not the asymmetric kernel as defined but rather its persistent closure. We want that $x\,R^{>}\,y$ and $y' \leq y$ and $x' \leq x$ then also $x'\,R^{>}\,y'$.

Proposition 2 *Let f be a monotone command-relation and $\mathfrak{T}$ a tree. The persistent closure of the asymmetric kernel of $f(\mathfrak{T})$ is both irreflexive and transitive.*

Proof. Let C denote the persistent closure of $R^{>}$. We then have $x\,C\,y$ iff there exist $x' \geq x$, $y' \geq y$ such that $x'\,R^{>}\,y'$. We have to show that never $x\,C\,x$ and that $x\,C\,y\,C\,z$ implies $x\,C\,z$. First, if $x\,C\,x$ then there exist $x', x'' \geq x$ such that $x'R^{>}x''$. But then $\neg(x' \circ x'')$, which cannot be. Second, assume $x\,Cy\,C\,z$. Then exist $x' \geq x$ and $y' \geq y$ such that $x'\,R^{>}\,y'$ and exist $y'' \geq y$ and $z'' \geq z$ such that $y''\,R^{>}\,z''$. This means that $x'\,R\,y'$ but not $y'\,R\,x'$ and that they do not overlap; that $y''\,R\,z''$, not $z''\,R\,y''$ and that they do not overlap. Likewise, we now have $f(x') \geq y'$ and $f(y'') \geq z''$ but neither $f(y') \geq x'$ nor $f(z'') \geq y''$. By local linearity we must have $y' \leq y''$ or $y'' \leq y'$. Assume the latter first. Then $f(y'') \leq f(y')$. Since $x' \nleq f(y')$ we have $x' \nleq f(y'')$. Suppose now that $f(z'') \geq x'$. Then since $z'' \leq f(y'')$ we also have $z'' \leq f(y')$ so that $f(z'') \geq y'$ means either $f(z'') \leq f(y')$ or $f(z'') \geq f(y')$. The first contradicts $f(z'') \geq x'$ the second means, however, $f(z'') \geq f(y'') \geq y''$, contradicting the choice of y'' and z''. So, then, $f(y') \leq f(y'')$. We can assume $f(y') < f(y'')$ from which follows $y' < y''$. There are two basic cases; either $f(y'') \ngeq x'$ or $f(y'') \geq x'$. Assume first $f(y'') \geq x'$. Then if $f(z'') \geq x'$, $f(z'')$ dominates the crosspoint of x' and z''. Since y'' does not overlap with z'' it does not dominate z''. Moreover, $f(z'') < f(y'')$. We know that the crosspoint of x' with y'' dominates the crosspoint of x' with y'. It cannot, however, dominate z'', because then y'' dominates z''. So, the crosspoint of x' with z'' dominates the crosspoint of x' with y'' and this shows that $f(z'') \geq y''$, which is excluded by choice of y'' and z''. Hence $f(z'') \ngeq x'$ and we are done. So this leaves the case $f(y'') \ngeq x'$. Then $f(z'') \ngeq x'$ as well, since we must have $f(z'') \geq f(y'')$. □

References

[1] Noam Chomsky. *Barriers*. MIT Press, Cambrigde (Mass.), 1986.

[2] Hubert Haider. Branching and Discharge. Technical Report 23, SFB 340, Universität Stuttgart, 1992.

[3] Richard S. Kayne. *The Antisymmetry of Syntax*. Number 25 in Linguistic Inquiry Monographs. MIT Press, 1994.

[4] Itziar Laka. Negation in syntax: on the nature of functional categories and projections. *International Journal of Basque Linguistics and Philology*, 25:65–138, 1991.

[5] Jan Wouter Zwart. *Dutch Syntax. A Minimalist Approach*. PhD thesis, University of Groningen, 1993.