4.1 Linear order in early generative grammar

It is interesting to note that linear order was not seen as an issue as such in mainstream generative grammar until the publication of Kayne (1994) and subsequent developments in the minimalist program, including Chomsky (1994, 1995: ch 4), which were in part a reaction on Kayne (1994). In the early phase of generative grammar when phrase structure rules were a component of the grammar, linear order was stipulated by the rules themselves: the rules generated strings. In a later phase, when phrase structure was generalised in the X-bar schema, word order was stipulated by word order parameter settings applied to universal X-bar structures. The way this was, and to some extent still is, conventionally represented is that the phrase structure tree, or the equivalent labelled bracket representation, encodes order as well as hierarchy. The tree (1) is trivially mapped onto the linear order to the right of the arrow, and the tree (2), which in hierarchical terms is equivalent to (1), onto the linear order to the right of its arrow.

(1)

```
NP
/ \NP
|   V  |
N1  N2
```

\( \Rightarrow \) \( N_1 \ V \ N_2 \)

(2)

```
V  \ NP
/   |
N1  NP
|   |
N2
```

\( \Rightarrow \) \( V \ N_2 \ N_1 \)
The mapping is trivial in that the linear order is derived simply by reading the terminal nodes in the tree from left to right. Thus phrase structure encodes hierarchy as well as linear order. Matters get more complicated when functional heads spelled out as affixes or clitics are taken into account, but theoretical devices such as head movement are designed in part to accommodate such phenomena. The fact that such a straightforward theory with its conventional representation of structure and word order seems to work, up to a point, making possible more or less adequate descriptions of a wide range of complex constructions in languages investigated, is one of the reasons for the success of generative grammar.

Consider what Chomsky (1965: 123-125) has to say on this topic: After observing that the phrase structure rules carry out two separate tasks, defining the syntactic relations and ordering of elements in deep structures, he says:

It has been suggested several times that these two functions of the categorical component be more sharply separated, and that the second, perhaps, be eliminated completely. Such is the import of the proposals regarding the nature of syntactic structure to be found in Curry (1961) and Šaumjan and Soboleva (1963). They propose, in essence, that in place of such rules as (69), the categorical component should contain the corresponding rules (70), where the element on the right is a set rather than a string:

\[
\begin{align*}
(69) \quad & S \rightarrow NP-VP \\
& VP \rightarrow V-NP \\
(70) \quad & S \rightarrow \{NP, VP\} \\
& VP \rightarrow \{V, NP\}
\end{align*}
\]

(Chomsky 1965: 124)

He continues (p. 125):
A priori, there is no way of determining which theory is correct; it is an entirely empirical question, and the evidence presently available is overwhelmingly in favour of concatenation-systems over set-systems, for the theory of the categorical component. In fact no proponent of a set-system has given any indication of how the abstract underlying unordered structures are converted into actual strings with surface structures. Hence the problem of giving empirical support to this theory has not yet been faced.

Since then, generative linguistic theory has undergone some quite radical changes, which will be sketched below. One result is that we now assume, within mainstream, minimalist-oriented generative grammar, that syntactic structure is constructed by Merge, in Chomsky’s (1993, 1995: ch. 4) sense, finally making phrase structure rules entirely superfluous. Another result is that we now have a theory for how to convert unordered structures into strings: the theory (or family of theories) based on Kayne’s (1994) Linear Correspondence Axiom.

In 1965 and up until the mid-seventies the formulation of transformations, the rules needed to derive passives, questions, relatives, etc. from the deep structure generated by the phrase structure rules, crucially depended on linear order. The phrase structure rules, generated strings of symbols (in fact, strings of words in the earliest versions, including Chomsky 1957). A transformation was a rule which took as input a string of symbols, a mix of terminal and nonterminal symbols and variables over symbols, and gave as output another string of symbols. Consider the English Dative Movement transformation, for example (adapted from Akmaijan & Heny 1975: 220):
This rule could apply to any phrase marker which met the structural description of the input. The structure (4a) does, for example, so it will be transformed into (4b):

(4)  a.  \([VP \ [V \ give] \ [NP \ a \ book] \ [PP \ to \ [NP \ Mary]]]\\)
    b.  give Mary a book

The terms of the rule are identified by linear order: They apply to an analyzed string, and they output another string. To apply to unordered structures (hierarchies of sets) the rules would have to be radically reformulated. As Chomsky put it: “a set-system such as (70) must be supplemented by two sets of rules. The first set will assign inherent order to the underlying unordered Phrase markers /.../. The second set of rules will be grammatical transformations, applying in sequence to generate surface structures in the familiar way.” (Chomsky1965: 125)

This formalism eventually became obsolete with the articulation of X’-theory and the structure preserving hypothesis, that is the realization that deep structures and surface structures (input and output of transformations) all conformed to the same principles of phrase structure (Emonds 1970,1976, Chomsky 1970, Jackendoff 1977). This meant that transformations could be formulated so as to apply to structurally defined constituents, moving them from one structurally defined position to another structurally defined position, or deleting them from structurally defined positions. Work by Chomsky (1973, 1976), building on Ross (1967), and by Rizzi (1990) further characterised the locality principles which movement and binding are subject to, making it possible to generalize and simplify the
rules themselves, to the point where they could eventually be subsumed under the maximally
general rule schema ‘Move α’ (Chomsky 1981, 1982).

This should make it possible, in principle, to shift to a ‘set-system’, where syntactic
relations (theta-role relations, scope, movement, binding, modification, case, agreement, etc.)
are defined in terms of hierarchic relations among constituents which form sets, not ordered
multiples. The attraction of such a system is, primarily, that it appears to be empirically
supported: Syntactic relations can typically be characterized precisely and exhaustively
without any reference to linear order. Linear order as such does not affect meaning, or put
more cautiously, does not add anything to aspects of meaning which are encoded at LF,
including theta-roles, scope relations, binding relations, definiteness, tense, mood, aspect, etc.,
and even (though much more arguably) relations such as topic and focus. What does affect,
and determine, those aspects of meaning is structure: dominance, c-command, and structurally
defined locality.2 There is a small residue of syntactic relations or constructions where linear
order has been claimed to be crucial (see Kayne 2013). However, if the large majority of
syntactic relations can be exhaustively characterized without reference to linear order, then it
is reasonable to assume that all of them can be, and that the few apparent counterexamples
have other explanations.

While this was widely accepted within generative linguistics in the seventies and
eighties, it did not lead to factoring linear order out of syntax until a decade or two later,
mainly because inclusion of ordering stipulations (implicit or explicit) in the syntax was seen
as innocuous, and because there was no credible alternative. As pointed out by Chomsky
(1965), dissociating linear order from hierarchy requires a theory of the mapping of hierarchic
structure onto linear order.
4.2 Word order parameters

To recapitulate, prior to X’-theory, the linear orders that were found in any given language depended on the phrase structure rules of the language and which transformations applied to the derived phrase-markers, where the transformations themselves were dependent on linear statements. Subsequent developments in the theory, including in particular X’-theory, made it possible to generalize and simplify the phrase structure rules, and to formulate transformations without reference to linear order.

Finally, with the generalization of X’-theory to all phrase types, including functional projections (Stowell 1982, Chomsky 1986), phrase structure rules became all but redundant, as they stipulated properties that followed, at least in part, from universal principles. Instead, the order of constituents in D-structure became a matter of setting word order-related parameters. This was made possible by the idea that syntactic variation across languages is a matter of a limited number of parameters, which were left open by UG, and were fixed in the course of language acquisition (Chomsky 1981). The order of constituents in the X-bar schema was one of the parameters. The order between specifier and X’ (including subject and predicate as a special case) seems to be almost universally specifier > X’. Assuming that it is universal (Whitman 2008: 234, 251), the only parameters would concern the order of head and complement and head and modifier. A highly attractive idea, which seemed to be supported by the findings of Greenberg (1966) and other research in Greenberg’s footsteps, including Dryer (1992), was that the head-complement (and head-adjunct) parameter settings were highly general; typically there would be one setting for nearly all types of heads within a language. The strongest position was that the setting is generalized to all heads: A complement precedes or follows the head, across all categories (see chapter 2). Apparent counterexamples would then be the result of movement, resulting from other parameter
settings which had the effect of triggering movement. For example, one idea explored in work by Travis (1984, 1994) and Koopman (1984) was that the direction of theta-role assignment and case-assignment were parameterized, independently of each other and independently of the head-complement parameter. A language could have, for example, head-complement order, but object case-assignment directed right-to-left. The result is that all the complements of a verb follow the verb, except DPs which have to precede the verb to be assigned Case (see Koopman 1984: 120-130). A weaker theory is that languages have a default head-complement direction setting for all categories, but there can be exceptions acquired on a case by case basis, the system being more marked the more exceptions there are.

However, even at this point the grammar can still be seen as embodying a set of generalised phrase structure rules: $\text{XP } \rightarrow \text{YP X’}$ (specifier-head), $\text{X’ } \rightarrow \text{X ZP}$ (head-complement), $\text{XP } \rightarrow \text{WP XP}$ (adjunction), all subject to order parameterisation, with the possible exception of the specifier-head rule. The notion ‘base order’ or ‘D-structure order’ therefore makes sense: it is the order of words stipulated by the language particular version of the phrase structure rules, once the terminal nodes have been substituted by lexical items (lexical insertion).

The D-structure is typically modified by movement of heads and/or phrases, deriving S-structure. With the possible exception of ‘affix hopping’ (see Lasnik 2000), movement is always upwards, but can be either leftwards or rightwards, in the sense that, when the structure is spelled out from left to right, the moved constituent either precedes or follows its trace/original position. It was well known that leftwards movement was considerably more common, and that rightwards movement, when it took place, was considerably more constrained (Ross 1967/1986; this was well known but not well understood).
At S-structure, finally, the linear order is read off the structural representation, spelling the terminal nodes out from left to right, before any movement applies which affects the structure without affecting PF (covert movement). This is part of the derivation of PF.

Thus, the linearization aspect of the mechanics of the system was basically not an issue, at this point (the eighties and early nineties). Word-order-related issues concerned mainly how to combine a restrictive theory of parameterised phrase structure with a restrictive theory of movement to account for observed cross-linguistic variation regarding word order.

4.3 Narrow syntax without linear order

The situation in current theory is quite different. Minimalist theorising now favours a dissociation of hierarchy and linear order. In the most radical view, which is arguably the mainstream view at present, the Narrow Syntax (the derivation of LF; henceforth narrow syntax) only recognises hierarchy. Its only structure-building operation is binary Merge in the sense of set-merge, combining categories successively pairwise, forming sets. The so formed sets are labelled, with the label equal to one of the merged categories (the head), but they have no order; the left-to-right order in the tree is just a notational convention forced upon us by the two-dimensional representation. The order only comes about in the derivation of PF, after spell-out (i.e. after the derivation splits into an LF and a PF branch). LF has no linear order.

This idea is closely connected with the radical idea that all linguistic variation is located in the derivation of PF. Narrow syntax, that is the derivation of LF by Merge, Agree, and whatever other operations are required, would be completely uniform; see Chomsky (2001), Burton-Roberts & Poole (2006), Burton-Roberts (2009), Berwick & Chomsky (2009), Boeckx (2011), Sigurðsson (2011a, 2011b) for different implementations of this idea. The idea is attractive for several reasons. To begin with, we know that there is a huge amount of
variation in PF: Most or all languages will have words for things like eyes and ears and activities like walking and eating, and so on, but they are not pronounced the same (in sign languages they are not pronounced at all, but signed). Furthermore, the system of vowels and consonants varies across languages, intonation patterns vary, a functional morpheme can be a free word in one language but a suffix or a prefix in another language, etc. etc. Word order is another obvious point of much cross-linguistic variation. If word order variation (SVO vs. SOV, Adj-N vs. N-Adj, etc.) is also located in PF, then it may well be that all variation is.\footnote{5}

We also know that there is not much variation that could be called semantic. It is, in fact, controversial whether there are any semantic parameters at all; see Chierchia (1998), von Fintel & Matthewson (2008), Ramchand & Svenonius (2008). If there aren’t, then this supports the contention that narrow syntax is uniform across languages. It should be noted that the works referred to above all argue that there are at least some semantic parameters (or, more generally, variation which is semantic in nature).

\subsection*{4.4 The LCA}

If Merge can only build hierarchically organized sets, and narrow syntax only recognises relations that are based on hierarchy, then, as pointed out by Chomsky (1965), we need a theory of the mapping of syntactic structure onto linear order. The most influential one proposed to date is the theory based on the Linear Correspondence Axiom (LCA), proposed by Kayne (1994), an important milestone in generative linguistics. (3) is a simplified formulation of the LCA:

\begin{quote}

\end{quote}
(3) Where \( \alpha, \beta, \) and \( \gamma \) are terminal elements (lexical items), \( \alpha \) precedes \( \beta \) if and only if \( \alpha \) asymmetrically c-commands \( \beta \), or \( \alpha \) is dominated by \( \gamma \), and \( \gamma \) asymmetrically c-commands \( \beta \).

Given that syntactic constituents must be assigned a linear order (to be pronounceable and parsable) every linguistic expression consisting of more than one pronounced item must observe the LCA.

In the context of a theory assuming that narrow syntax does not recognize linear order, the LCA thus says that asymmetric c-command in narrow syntax maps onto precedence, hence linear order, in PF.\(^6\) For Kayne, the motivation for the LCA was not to factor linear order out of syntax, though. In fact, Kayne explicitly distances himself from that view (Kayne 1994: 48ff., 2013). Instead, the LCA was proposed as an explanation of the pervasive left-right asymmetry of syntax, and in order to derive, instead of stipulating, the properties of X-bar theory. The asymmetries include: almost all specifiers are on the left (including subjects), almost all movement is leftwards, almost all head-adjunction is on the left, and head-final order is not the mirror image of head-initial order but is more constrained than head-initial order. Another left-right asymmetry which the LCA can explain is Greenberg’s Universal 20, discussed in chapter 7. The LCA entails that these asymmetries are universal and exceptionless, and any counterexamples only apparent. The reason why specifiers are (always) on the left would be that specifiers, by definition (see note 6), asymmetrically c-command, hence precede, the head they are specifiers of. The reason why movement is (always) leftwards would be that movement is always upwards (cf. the extension condition in Chomsky 1993, Freidin 1999), hence the moved constituent asymmetrically c-commands its trace or copy and the sister of the trace or copy, hence it precedes the trace or copy, and in this sense movement is leftwards.
Furthermore, Kayne (1994) notes that the LCA provides an explanation of X-bar theory: It follows from the LCA that a phrase cannot consist of two $X^0$ heads, or of two maximal XPs: Both types of phrases would be symmetric, and therefore could not be linearized; linear order presupposes asymmetric c-command. That is to say, it follows from the LCA that phrases must have one and only one head; endocentricity (hence labelling of phrases) is thus a consequence of the LCA; see, however, Guimarães 2008.

Kayne (1994) argues that this means that the LCA, and hence linear order, must be a property not just of PF, but of every syntactic representation: “It follows that to declare the LCA inapplicable at some level of representation – say LF – would be to declare inapplicable to that level of representation all the restrictions familiar from X-bar theory (existence of at least one and at most one head per phrase, etc.).” (Kayne 1994: 49) This does not follow, though, if the successful derivation of a linguistic expression requires that it converges, in the sense of Chomsky (1995: ch. 4), i.e. it must satisfy the output conditions at PF and LF, the two interface levels. If a derivation crashes at one of the interfaces the derivation is ruled out. So if, for example, a small clause [DP AP] crashes at PF because it is too symmetrical to be assigned a linear order (see Moro 2000), the derivation is thereby cancelled, i.e. the expression is underivable. This means that narrow syntax must provide the tools and building blocks required for phrases to observe the LCA at PF, even if these tools and building blocks would serve no purpose at LF. Phrases must have a structure such that they are linearizable at PF, even though the linearization is activated only after spell-out.7

One of the most notorious consequences of the LCA is that the underlying order of every phrase is spec > head > complement. Take the case of the TP in (4):
The spec-head-complement order follows since DP1 asymmetrically c-commands T, hence its daughter D and any terminal daughters of NP must precede T, while T asymmetrically c-commands V and DP2, hence must precede them. It follows that any other order of spec, head, and complement must be the result of movement. For example, the order spec > complement > head, where it occurs, must be the result of movement of the complement to a position where it asymmetrically c-commands the head. “The LCA entails that all languages are underlyingly SVO languages” is a popular interpretation of this consequence of the theory. This consequence of the LCA has been the subject of much debate, and is probably the one property of the theory which has caused most resistance (see Svenonius 2000).

In a theory like that of Chomsky (1995: ch. 4) where linear order is factored out of syntax, what does it mean to say that the underlying order of every phrase is spec> head > complement? Can there be an ‘underlying order’? What it means is that, unless something happens to the phrase, such as movement of the complement to a position higher than the head, or movement of the head to a position higher than the specifier, the phrase will be spelled out as spec > head > complement. In narrow syntax the phrase is the result of Merge; first merge of head X’ and a phrase YP forming a set \{X’, YP\}, which is assigned the label X (call it X’); then merge of another phrase ZP, forming a larger set \{ZP \{X’, YP\}\} again assigned the label X (call it XP), etc. Constituents can be moved, i.e. can be merged again (by internal merge), and can enter into various syntactic relations (Agree, binding, etc.) subject to locality conditions which only recognize structure. The LCA, as regarded here (not as regarded by Kayne 1994, 2013), is activated only at spell-out, determining unequivocally
the linear order of words and morphemes making up the terminal nodes of the hierarchical structures constructed by the narrow syntax.

It could be noted that the LCA-based theory of word order variation is not easily reconciled with the idea that all linguistic variation is a matter of externalisation. This idea would seem to presuppose that all movement which creates chains and affects scope, i.e. which has effects at LF, is universal, the cross-linguistic variation being a matter of which copies are pronounced (Groat and O’Neil 2000, Bobaljik 2002). For example, so called wh-in-situ would be a matter of pronouncing the lower copy in a wh-chain, in a theory where wh-movement is universal. Any movement which does not affect interpretation, including the movement of a complement around a head just to derive OV order (assuming the LCA), would then be postsyntactic (applying in ‘the PF-component’). But then, if it is the case that OV order feeds operations which do have an effect on interpretation, there is a problem. For example, it is well known that OV order in the VP is conducive to scrambling, a movement rule which has interpretive effects, in particular on definiteness, and on information structure more generally (see chapter 10 for some discussion). This is not easily reconciled with the hypothesis that the derivation of OV order is postsyntactic (but see Richards 2004, Fox and Pesetsky 2005) for PF approaches to these problems).

4.5 The bottom pair problem

Incorporating the LCA in Bare Phrase Structure (BPS) theory (Chomsky 1994, 1995) leads to a problem with regard to the ‘bottom pair’, i.e. the two X⁰ constituents merged first in any syntactic derivation. The problem is that they are, by definition, two non-branching sister categories, and as such will not be assigned a linear order by the LCA. For example, the tree
resulting from merging a definite article with a noun will look like (5) (the tree here representing a set consisting of the D the and the noun book labelled by the head D).

(5)

\[
\begin{array}{c}
D \\
| \downarrow \\
N
\end{array}
\]

the book

In terms of c-command, this is a symmetrical structure, hence not linearizable. Any subsequently merged constituents will asymmetrically c-command the article and the noun, and therefore can be linearized in relation to them, but the two can’t be linearized in relation to each other.

The problem did not arise in Kayne (1994), because (a) he does not adopt the BPS principle that Merge is the only structure-building operation, and therefore does not rule out non-branching nodes, and (b) he adopts a definition of c-command based on first node, not first branching node. Thus (5) can be analyzed as (6) in Kayne (1994), where D asymmetrically c-commands N, so that the will precede book when spelled out.\(^9\)

(6)

\[
\begin{array}{c}
DP \\
| \downarrow \\
NP \\
| \downarrow \\
N
\end{array}
\]

the book

Various solutions to the bottom-pair problem have been proposed in the literature. Zwart (2011a) argues for a phrase structure theory which differs from the one proposed by Chomsky in that merge does not derive sets, but ordered pairs: merge is always an asymmetric process,
not ‘merging a and b’, but always ‘merging a with b’, which avoids the bottom pair problem – by reinstating linear order in narrow syntax.

Another possibility is that a structure like (5), merge of a functional X° and a lexical X°, is never the bottom pair, because the lexical X° has the structure \([f, \text{root}]\), where \(f\) encodes lexical/syntactic category, following Marantz (1997) and Borer (2005?). If \(f\) is \(n\), for example, merging it with the root BOOK yields the noun \textit{book}. This has the consequence that an article merged with \([n, \text{BOOK}]\) will asymmetrically c-command the constituents of \([n, \text{BOOK}]\), and therefore, given the LCA, will be spelled out preceding these constituents, unless they undergo movement, which will yield a postnominal article. Of course, this leaves the problem how \(f\) and the root are linearized. I will return to this issue in the next section. Chomsky (2012), following the “Marantz-Borer conjecture”, takes this line: the bottom pair is \([f, \text{root}]\).

Chomsky (2012) is not concerned with linear order, though, but with projection and labelling: How is the label determined in the bottom pair where, apparently, two X° constituents are merged? He proposes that \([\text{root}]\) does not qualify as a label, and therefore \([f, \text{root}]\) will always be labelled \(f\).

### 4.6 Alternatives to the LCA

Obviously, if the observed asymmetries that the LCA is based on are not universal, but at most tendencies, then alternatives have to be considered. One alternative is articulated by Abels and Neeleman (2009). They argue that the order of heads, complements, and specifiers is parametrized, in the sense that heads, complements, and specifiers can be externally merged (base-generated) in any order, fixed independently for each language. They discuss in particular Greenberg’s Universal 20, a generalization concerning word order in noun phrases: see chapter 7. However, they argue, in line with Kayne (1994), that movement (internal
merge) is always leftwards, which explains a number of asymmetries which, arguably, are universal, such as the absence of ‘verb-second-from-the-end languages’ or rightwards wh-movement languages. They also suggest that the ban on rightwards movement is ultimately an effect of constraints on processing of syntactic dependencies. (see chapter 5 for discussion).

Another alternative is articulated by Haider (2000, 2005, 2012). The core idea in Haider’s theory is that given a highly restrictive, yet reasonably simple characterization of the relation between structure and word order, allowing for the variation observed, a model where “head-final order is more basic than head-initial order” (Haider 2000) gives a better account of the empirical facts when comparing head-final and head-initial systems. The theory is based on the universal Branching Constraint (7) and the parametrized direction of licensing of dependents by a head (8).

(7) The Branching Constraint: Projection-internal branching nodes on the (extended) projection line follow their sister node.

(8) Licensing to the left gives head-final order.
   Licensing to the right gives head-initial order.

The projection of, say, a VP headed by a put-type verb will have the following structure in an OV language, that is a language with licensing to the left.

(9) \[
\begin{array}{c}
\text{VP} \\
\text{NP} \\
\text{NP} \\
\text{PP} \\
\text{V} \\
\end{array}
\]
The ‘projection-internal branching nodes’ here are the two V’-nodes, which follow their sister, i.e. the specifier. Here, because the bottom branching node is head-final, the licensing of the dependents is ‘harmonic’ with the branching direction: All the heads (in the X-bar sense, that is X° and X’) license their dependents leftwards. The VP can then be the complement of, for example, an auxiliary, forming another head-final phrase [VP Aux]. In a VO language the structure will be (10).

(10)

Here, the licensing direction is disharmonic with the branching direction: In order for the head to license the dependents it has to move, deriving a VP-shell structure. That is to say, head-initial order is always derived by head-movement – in contrast with the LCA-based theory, where head-final order is derived by complement movement. Haider provides arguments, in particular from the syntax of predicative modifiers in favour of this theory.10

A related theory is articulated by Emonds (2013). As pointed out by Emonds, if head-final order is the universal default, and if the X-bar-theoretic notion that X’-level phrases are heads is adopted, then the fact that specifiers (apparently) always precede their sister X’ would be a consequence of the universal head-final order. That is to say, one of the most powerful empirical arguments for the LCA (specifiers always precede the head) can be construed as an argument in favour of the ‘opposite view’, that head-final order is the universal ‘basic order’.
Again we may ask, in a model where linearization is factored out of narrow syntax, what does it mean to say that head-final order is the basic order? The answer is essentially the same as in relation to the LCA in section 4.4: Unless there is movement, the derived structure is spelled out as head-final. Embedding Haider’s (2000) theory in such a model, with the crucial notion ‘licensing direction’, ‘the head $\alpha$ licenses $\beta$ to the right’ (the marked option) requires that the head $\alpha$ moves to a position where it c-commands $\beta$. When spelled out, this yields $\alpha \triangleright \beta$. Licensing on the right does not require any head-movement, so if no head-movement occurs in the syntax or after (cf. Chomsky’s (1995: ch. 4) suggestion that head-movement is postsyntactic), the structure is spelled out as head-final in PF.

4.7 A challenge for linearization theory: FOFC

This book is about a ‘word order condition’, involving the relation between linear order and structure, the Final-over-Final Condition (FOFC) (see chapter 1).

(11) The Final-over-Final Condition (FOFC)

A head-final phrase $\alpha P$ cannot immediately dominate a head-initial phrase $\beta P$ where $\alpha$ and $\beta$ are heads in the same extended projection.

What are the implications of this constraint for theories of linearization? We can note, first of all, that it cannot be stated just in terms of linear ordering specifications (directionality parameters), even if they are relativized to category (with different specifications for different categories). The issue does not arise in a language which is consistent head-initial or consistently head-final, but only in mixed order languages, that is languages where a given
head may either precede or follow its complement. But, as shown in chapter 2, a language with mixed word order may well allow, for example, the linear order V>O and the linear order VP>Aux, but still reject the combination of the two. This is the case, for example, in Finnish (see Introduction and particularly Chapter 2 for more detailed discussion of this example).

(12)  

(a)  
Anne on ostanut auton.  
[V>O]  
Anne has bought car  
‘Anne has bought a car.’

(b)  
Milloin Anne tullut on?  
[VP>Aux]  
when Anne come has  
‘When did Anne come?’

(c)  
*Milloin Anne ostanut auton on?  
[[V>VP O] AUX]  
when Anne bought car has

See Biberauer, Holmberg and Roberts (2014) for similar data from the Kaaps dialect of Afrikaans and Basque.

The LCA has consequences for FOFC, though. To begin with, the LCA excludes a derivation where the structure (13a) is linearized as (13b).

(13)  

(a)  
[AuxP [VP V O] Aux]  

(b)  
V>O>Aux

Since Aux asymmetrically c-commands V and O in (13a) it will precede V and O according to the LCA. Instead, from the point of view of the LCA (3), the structure of the string (13b)
could, *a priori*, be (14a) or (14b) (where F is a phonetically null), with VP moved to a position where it asymmetrically c-commands Aux.

(14) a.  

```
  VP  
    V  O  Aux  <VP>
```

b.  

```
  VP  
    V  O  F  Aux  <VP>
```

Since V and O are dominated by VP, which now asymmetrically c-commands Aux, V and O will precede Aux, by the LCA (3). However, this configuration is ruled out by FOFC formulated as in (11). In (14a) it will be ruled out where \( \alpha \) is Aux, a head in the extended projection of V (Grimshaw 1991, 2001). In (14b) it will be ruled out where \( \alpha \) is F, a head in the extended projection of V. The question is why (14a,b) are ruled out.

In the following I will review four theories all of which assume some version of the LCA, and all of which profess to explain FOFC in terms of some principle(s) of linearization. The first one is in Holmberg (2000a), the first paper addressing FOFC as a universal condition on the mapping of structure to word order. The second one is in Biberauer, Holmberg and Roberts (2014), the third one is in Sheehan (2011, 2013a,b), while the fourth one is presented here for the first time.

### 4.8 Holmberg (2000a)

This paper contains the original formulation of FOFC (without calling it that):

(15) If a phrase \( \alpha \) is head-initial, then a phrase \( \beta \) which immediately dominates \( \alpha \) is head-initial. If \( \alpha \) is head-final, then \( \beta \) can be head-initial or head-final.
In other words, seen from the point of view of the dominating phrase, head-final can only dominate head-final: hence we now call it the Final-over-Final condition. The paper focuses on Finnish sentential word order in particular, but demonstrates how the condition is valid also in other languages and other types of phrases. Sentential word order in Finnish is a telling example, though, because, although the unmarked order in Finnish sentences is uncontroversially head-initial, head-final order occurs as well, systematically in certain syntactic contexts, O preceding V, and VP preceding Aux. The syntactic context where the head-final order occurs is, primarily, when some constituent is fronted to initial focus position, for instance a wh-phrase in a question (see Chapter 10). We thus find Aux either preceding VP or following VP, and we find V either preceding its complement O or following it. Given that the marked head-final alternative is always optional, we would expect to find all the possible combinations of head-initial and head-final VP and AuxP occurring, given the right syntactic conditions. However, one combination does not occur, and is judged unacceptable: A head-initial VP dominated by a head-final AuxP (see Chapters 2 and 10 for other examples).


when Jussi would.have novel written

when Jussi novel written would.have

when Jussi written novel would-have
The explanation of FOFC in Holmberg (2000a) is based on the following theoretical premises:

First, based on a proposal by Svenonius (1994), it is assumed that selection, specifically c-selection (or ‘strict subcategorisation’, in the terms of Chomsky 1965), is always accompanied by movement. More precisely, c-selection is formally movement, or we might call it incorporation, of the selected feature into the selecting head. The rationale is that the c-selection is a strictly local phenomenon. This feature-movement (or incorporation) has three different realizations, subject to parametric variation (see also Julien 2002). Consider the tree (17), where F c-selects G:

(17)  
```
       F
      / \  
     /   \ 
    G    XP
```

(a) G can move by pure feature movement, which is always covert.

The resulting structure is spelled out as F>XP>G.

(b) G can move by head movement (overt incorporation of G in F).

The resulting structure is spelled out as G>F>XP (typically with F a suffix on G).

(c) G can move pied piping GP.

The resulting structure, a harmonic head-final structure, is spelled out as XP>G>F.

Now assume that the tree is (18):
(a) G can move by pure feature movement, which is always covert. The resulting structure is spelled out as $F \rhd G \rhd XP$.

(b) G can move by head movement (overt incorporation of G in F). The resulting structure is spelled out as $G \rhd F \rhd XP$ (typically with F a suffix on G); same as in (17).

(c) G can move pied piping GP.

This will yield the FOFC-violating structure (19).

\[
\begin{array}{c}
\text{(19)} \\
\begin{array}{c}
\text{FP} \\
\text{GP} \\
\text{G} \quad \text{XP} \\
\text{F} <\text{GP}>
\end{array}
\end{array}
\]

The theoretical device which rules this structure out is (20), taken to be a universal condition on c-selection (this is a simplification; see Holmberg (2000a) for a more detailed formalisation); see Richards (2010) for a more recent formalisation and contextualisation of this general idea, which is also applied to FOFC by Richards.

\[
\begin{array}{c}
\text{(20)} \\
\text{C-selection requires adjacency between selector and selectee.}
\end{array}
\]
Feature-movement and head-movement will always yield adjacency. Phrasal movement will yield adjacency only if the pied-piped phrase is head-final.

While in the case of the options (a) and (b), the adjacency is ensured in the (narrow) syntax, by incorporation of the selected head in the selecting head, overtly or covertly, in the case of (19), it is clearly a PF-matter: a spelled out phrase will intervene between selector and selectee. In the case of (17) Holmberg (2000a) assumes that GP is itself derived by phrasal movement of XP, as required if the LCA is assumed, which it is in Holmberg (2000a). It will thus have the structure (21), after movement of XP.

\[
\begin{array}{c}
\text{FP} \\
\text{GP} \\
\text{XP} \\
\text{G} \quad <\text{XP}> \\
\end{array}
\]

By hypothesis the copy of XP does not intervene between G and F for the purposes of (20). That is to say, the required adjacency is a PF matter, a morphological issue.

The argumentation in favour of the principle (20) in Holmberg (2000a) begins with the observation that in the case where F is an affix (for instance a suffixal tense head), adjacency is required because otherwise an ‘affix-feature’ (a morphological matter) remains unchecked. It proceeds to the suggestion that the affix-feature is just a special case of the c-selection feature. It is, as it were, just a morphological accident that in some cases the selecting head is an affix, i.e. is spelled out as a bound morpheme, and sometimes a free morpheme: adjacency is a requirement in either case.

It is a well known generalization that the heads in a head-final tree always line up adjacent to each other, while the heads in a head-initial tree often don’t. FOFC has this effect:
no complement can intervene between the heads in a head-final structure. But it is a more general condition, since typically adjuncts cannot intervene between the heads in head-final tree, either (as mentioned in Chapter 2 and further discussed in Chapter 5). I illustrate this with the following Finnish construction (based on Holmberg 2000a). (22) is the unmarked harmonic head-initial structure, and the adverb can be inserted in a variety of places, including between the auxiliary and the verb.

(22) Ei Jussi ole (vieläkään) hyväksynyt (vieläkään) sitä ajatusta (vieläkään). [Finnish]
not Jussi has yet accepted yet that proposal yet
‘Jussi still hasn’t accepted that idea.’

(23) is the head-final version. Again the adverb can be inserted in more than one place, but not between the verb and the auxiliary.

(23) Ei Jussi (vieläkään) sitä ajatusta (vieläkään) hyväksynyt (*vieläkään) ole (vieläkään).
not Jussi yet that proposal yet accepted yet has yet
‘Jussi will surely not accept that proposal either.’

This is explained by (20). Consider the relevant part of the derivation of (23b) (simplified for ease of exposition): In (24), the object has moved to spec,VP, and the auxiliary has been merged, c-selecting the verb meaning ‘accepted’.

(24) \[
\begin{array}{c}
[\text{AuxP } \text{ole } \text{[VP } \text{[DP } \text{sitä ajatusta } \text{]} \text{[V' hyväksynyt } \text{<DP> } \text{]}]]
\end{array}
\] have that idea accepted
Now assume that prior to movement of V pied piping the VP, the adverb is merged as an adjunct to AuxP, deriving (25a). Only then the VP moves, deriving (25b).

(25) a. \[\text{AuxP vieläkään [AuxP ole [VP [DP sitä ehdotusta ] [v’ hyväksynyt <DP> ]]]} \]

\[
\text{yet have that proposal accepted}
\]

b. \[\text{AuxP [VP [DP sitä ehdotusta ] [v’ hyväksynyt <DP> ] [AuxP vieläkään [AuxP ole <VP> ]]]} \]

\[
\text{that proposal accepted yet have}
\]

This will yield the ungrammatical version of (23b). Principle (20), the principle explaining FOFC in Holmberg (2000a), will rule it out.

(20) is still dubious as a principle of UG, though. C-selection is a narrow-syntactic property (or operation), notoriously sensitive to features visible at the LF-interface such as [±WH] or [±Finite], and notoriously insensitive to morphological features like whether a complement is actually pronounced or not (for instance, the VP complement of Aux can be deleted in some contexts, in many languages). How can such a property (or operation) be subject to a PF-adjacency condition? If we take this proposal seriously, it would seem to have radical architectural consequences. See Richards (2010), though, for a theory which does take that step.

The alternative is that the adjacency is not in any sense driving the derivation (in a derivational model), or serving as an output condition (in a representational model), but is just an incidental consequence of the how the derivation of head-final structure operates.12
4.9    FOFC in Biberauer, Holmberg, and Roberts (2014) (BHR)

The LCA is a crucial component in the theory articulated in Biberauer, Holmberg and Roberts (2014) (henceforth BHR), as well. First, assuming the LCA, any string where (the constituents of) VP precede Aux, for instance the well formed substring *ein Auto gekauft hat* in (15), must be derived by movement. The structure of (26a) will be roughly (26b).

\[(26)\]  
\[a. \quad \ldots \text{dass Anne ein Auto gekauft hat.} \quad \text{[German]} \]

\[b. \]
\[
\begin{aligned}
\text{VP} \\
\text{O} \quad \text{V} \quad \text{Aux} \quad <\text{VP}> \\
\end{aligned}
\]

BHR explicitly assume that the relevant movement is movement of the sister of $X^\circ$ to spec,XP, ignoring the more complex alternative where there is an additional abstract head above XP attracting the complement of $X^\circ$ to it’s spec (see footnote 10)), and ignoring the arguments in the literature against such very local movement.$^{13}$ Assuming the LCA, the order $O>V$ in the VP in (26b) must also be derived by movement, i.e the structure of the VP is, roughly, $[VP \ [VP \ V \ <O>]]$, as in (21a). In these terms, the contrast between the licit structure (21a) and the illicit structure (21b) suggests the following generalization: a VP consisting of V and O cannot move around Aux unless O has moved around V (deriving $O>V$) (where the expression ‘YP moves around $X^\circ$’ is shorthand for ‘the sister YP of $X^\circ$ moves to spec,XP’).$^{14}$ More generally, the movement required to derive an extended projection with head-final order must ‘start at the bottom’, the sister of the lexical head L moving around L to a position where it asymmetrically c-commands L, before the phrase headed by L can move around the next
head merged with LP, in the extended projection of L (why the notion ‘extended projection’ is crucial here will be made clear below). Complex head-final structures are derived by iterated movement of successively larger complements, so called snowball movement (Kayne 1994, Jayaseelan 2009, Cinque 2005a), but always starting at the bottom of the tree. Furthermore, the iterated movement of complements must be strictly successive: it cannot ‘skip’ a head, or a FOFC-violation will ensue. Consider the following example from Finnish:

indeed I notes read could have  
‘I have actually been able to read music.’

indeed I read notes could have

c. *Kyllä minä osannut nuotteja lukea olen. [AuxP [vP v [VP O V ]] Aux]  
indeed I could notes read have

The extended projection of the verb here includes the restructuring verb osata ‘know how to’, labelled v, and the auxiliary olla ‘have/be’. In the grammatical (27a) the object has moved around the verb (being spelled out as O>V), the VP has moved around the v, and the vP has moved around the auxiliary. In (27b), O has not moved around V, but VP has moved around v, and vP around T. The result is ungrammatical, a FOFC-violation. In (27c), O is moved around V, then VP is not moved around v, but vP is moved around T. The result is ungrammatical, again violating FOFC. This is not because the FOFC-violating restructuring
verb must follow its complement as an inherent property: In other contexts the verb *osata* can precede its complement:

(28) Kyllä minä olen osannut  nuotteja lukea.  \[ {\text{AuxP} \; \text{Aux} \; {\text{vP} \; {\text{vP} \; {\text{O} \; \text{V}}}}]} \]

indeed I have been.able.to notes read

‘I have actually been able to read music.’

BHR propose, following Chomsky (2000, 2001) and much subsequent work, that movement is triggered by a feature on the head which is the target of the movement. Chomsky refers to the feature as the EPP-feature (Chomsky 2000), or OCC (Chomsky 2001), or the edge-feature (Chomsky 2008). BHR use the symbol ^ (caret) for the movement-triggering feature. This feature has different instantiations depending on which syntactic feature(s) it is associated with. Thus A-movement is triggered by ^ coupled with a phi-feature probe on T or v (i.e. a set of unvalued phi-features needing a matching valued set to assign values to it; Chomsky 2001). A-bar movement is triggered by ^ associated with a head in the C-domain. The movement of the sister YP of X° around X°, which has head-final order YP> X° as its principal effect, is triggered by ^ associated with the categorial feature [±V] of X°. This feature is introduced in a derivation with the lexical head. This is why the movement deriving head-final order starts at the bottom. Adapting the idea that the bottom pair is the merge of a root and a categorial feature (see section 4.5 above), ^ will be a property of the categorial feature. It is a parametrized property of the categorial feature; typically a given lexical category, e.g. P, in language L has ^ consistently or lacks ^ consistently (the language is prepositional or postpositional). It is also possible, though, for a class of P in L to have ^, while the complementary class lacks ^, as in Finnish, where most adpositions are postpositions but there is a sizeable minority of prepositions (Manninen 2003; see Chapter 1). There are also
languages where ^ can be an optional accompaniment of a given lexical category. Finnish is, again, an example. As shown in (12) and (16), and as will be discussed in more detail in Chapter 11, VP, vP and AuxP can be head-initial or head-final in Finnish.

An extended projection, in roughly Grimshaw’s (1991, 2001) sense, is defined by the categorial feature of the lexical head. The extended projection of V is the sequence of functional heads sharing the categorial feature [+V] emanating from the verb, including at least v, Aspect, T, modals (in languages where they are verbal), and arguably C. The extended projection of N is the sequence of functional heads sharing the categorial feature [–V], emanating from the noun, including number, noun class, quantification, D, and arguably P. BHR now propose that the movement trigger ^ ‘spreads’ or is copied from head to head along the spine of the extended projection. In (29), representing the structure of the bottom part of a head-final transitive sentence, for example (27a), ^ is introduced in the derivation as an accompaniment of [+V]. It triggers movement of the object DP to spec,VP.

\[(29)\]

\[
\begin{array}{c}
\text{v} \\
\text{ [+V]} \\
\text{DP}
\end{array}
\]

\[
\begin{array}{c}
\text{VP} \\
\text{<DP>}
\end{array}
\]

Subsequently the transitivising light verb v is merged. The light verb is marked as c-selecting [+V] (marked by a strict subcategorization feature), a condition for successful merger with VP. BHR propose that v is not itself inherently valued [+V], but ‘inherits’ or copies this feature from its selected complement. In fact, the categorial feature copying can be seen as a necessary consequence of the selection relation: the selection feature is ‘checked’ by copying the categorial feature of the complement (see Svenonius 1994, Holmberg 2000a, Julien...
The result is (30), where ^ is copied by v, triggering movement of VP around v.

Following BHR, the presentation here ignores the subject, also merged with vP.

The movement trigger ^ can be, but need not be, copied along with the categorial feature it accompanies. This would be a point where UG is underspecified, thus allowing variation. And variation is what we find: in some languages ^ is copied all the way up to the highest head of the extended projection of L. This yields a harmonically head-final extended projection of L. In other languages it stops at some intermediate head along the projection line from L to the highest head in the extended projection of L. The stopping point is subject to parametric variation. Thus we find SOV languages such as (31a) (^ copying all the way to C), (31b) (^ copying stops at T), and (31c) (^ copying stops at V).

(31) a. O>V>T>C (Japanese)
    b. C>O>V>T (Hindi)
    c. C>T>O>V (Dutch)

Now if copying of ^ is a consequence of selection, it follows that it is strictly local (a head-to-head relation). This would then be what rules out (27c). To derive (27c), repeated here as (32a), ^ would have to be copied by Aux while skipping v, as shown in (32b):
(32) a. *Kyllä minä osannut nuotteja lukea olen. [Finnish]
    
    indeed I could notes read have

b. \[[\text{AuxP} \ [\text{vp} \ [\text{vp} O V ] ] \text{Aux} ]\]
   
   \[+[V] \ [+V,^\] \ [+V,^]\]

This, BHR argue, is ruled out by locality: Aux cannot c-select V, as required for feature copying. It can only c-select v, but v in (32b), although it has the right categorial feature [+V], does not have ^. Hence Aux cannot have ^ triggering L-movement, hence (32a) cannot be derived by the system. BHR formulate the locality condition on selection in terms of Rizzi’s (1990) Relativized Mini-mality:

(33) **Relativized Mini-mality** (adapted from Rizzi 2001):

In a configuration X...Y...Z where X asymmetrically c-commands Z, no syntactic relation R can hold between X and Z if Y asymmetrically c-commands Z but does not c-command X, and R potentially holds between X and Y.

The following is the reason why the notion ‘extended projection of L’ is crucial: BHR observe that FOFC appears not to apply to nominal complements to verbs, exemplified by the following two German examples.

(34) a. Johann hat \[[\text{vp} \ [\text{dp} \text{ einen Mann} ] \text{ gesehen} ]\].
    
    John has a man seen
    
    ‘John has seen a man’

John is to Berlin gone

‘John has gone to Berlin’

Here, by assumption, the complement of the main verb, a DP in (34a), a PP in (34b), is moved around V, deriving a head-final VP. But the DP and the PP are head-initial, without any resulting FOFC-violation. Consider again the formulation of FOFC in (11), repeated here:

(11) **The Final-over-Final Condition (FOFC)**

A head-final phrase \( \alpha P \) cannot dominate a head-initial phrase \( \beta P \) where \( \alpha \) and \( \beta \) are heads in the same extended projection.

The difference between these cases and those of (12c) and (16b,c) is that \( \alpha \) and \( \beta \) in (34a,b) are not heads in the same extended projection, while they are in (12c) and (16b,c). In (34a,b) \( \alpha \) is [+V], the head of VP, while \( \beta \) is [–V] (the head of DP in one case, the head of PP in the other).

BHR present arguments that the crucial difference really is categorial feature value. In particular they reject the idea that the crucial difference between (34a,b) and (12c)/(16b,c) is that the moved (pre-head) complement is a referential expression, an argument of the verb, in (34b,c) but not in (12c)/(16b,c). That this is not a crucial distinction is shown (a) by the fact that sentential complements to V are subject to FOFC (see Chapter 1, section 1.3), and (b) that predicative noun phrase complements of V are not subject to FOFC. Consider the difference between (35a) and (35b).
(35) a. *...dass Johann niemals [CP dass er eigentlich ein angenommenes Kind sei] besprochen hat.*

That Johann never that he actually an adopted child be.SUBJ discussed has

b. ...dass Johann niemals besprochen hat [CP dass er eigentlich ein angenommenes Kind sei].

Adopted child be.SUBJ

‘...that Johann has never discussed the fact that he is actually an adopted child.’

The embedded clause in (35a) is an argument of the verb. When it precedes the verb, the result is ungrammatical because it violates FOFC, according to BHR (see also Chapter 2). When it is postposed to the verb the result is grammatical. This follows if (a) FOFC only applies internally to an extended projection, and (b) CP is a projection of [+V], part of the extended projection of the verb. An alternative hypothesis is that the reason why a CP complement must be postposed is that it is too ‘heavy’ to be preposed. This hypothesis is falsified by (36), where the complement of V is even more complex, yet can be preposed.

(36) ...dass Johann niemals [DP der Verdacht [CP dass er eigentlich ein angenommenes Kind sei]] besprochen hat.

That Johann never the suspicion that he actually an adopted child be.SUBJ discussed has

‘...that Johann has never discussed the suspicion that he is actually an adopted child.’
Under BHR’s theory, (36) is well-formed, in contrast with (35a) because the preposed complement in (35) is a DP, not a projection of [+V].

(37), finally, is an example of a nominal complement preposed to the verb which is not an argument, but a predicate.

(37) ... dat Johan [NP minister van buitelandse sake] geword het. [Afrikaans]

that Johan minister of foreign affairs become has

‘...that Johan has become minister of foreign affairs.’

If BHR are right, FOFC is a consequence of how selection, feature projection, and movement work, in the structure-building process. These are core processes in narrow syntax. The relation that FOFC has to linearization, in a model where linear order is factored out of narrow syntax, would be that the structure which would yield the ‘FOFC-violating’ word orders in (12c), (16b,c), and (35a) cannot be derived by the narrow syntax.

Summarizing, in BHR FOFC is a result of the following syntactic conditions:

(38) (a) Head-finality is a consequence of the movement-trigger \(^\wedge\) being paired with the categorial feature \([\pm V]\), which enters the derivation with the head of the extended projection.

(b) The movement-trigger \(^\wedge\) can spread with \([\pm V]\) from head to head along the spine of the extended projection, subject to parametric variation.

(c) C-selection relations are subject to Relativized Minimality.\(^{17}\)
Another model providing a principled explanation of FOFC is articulated in Sheehan (2011, 2013a,b). Like BHR, Sheehan assumes a model where linear order is factored out of (narrow) syntax. Unlike BHR, Sheehan assumes that heads are marked for ‘direction’ as part of their inherent c-selection features. For instance, in a VO language V is marked, as an inherent property, to precede the category it selects when spelled out at PF, formally notated as Vₚ, while in an OV language, it is marked to follow the category it selects when spelled out at PF, formally Vₖ.

Another distinctive property of the model is that every terminal category must be linearized relative to every other terminal category within a certain domain without relying on dominance. Thus, for example in the structure [Aux [ O V ]] there are three linear relations, three linear pairs: O-V, Aux-V, and Aux-O. Each of them must be unambiguously determined, though transitivity can be used to determine the order between heads which do not stand in a selection relation. In a structure [Neg [Aux [ O V ]]] there are six linear pairs which must be unambiguously determined: O-V, Aux-V, Aux-O, Neg-Aux, Neg-V, Neg-O. And so on, for more complex structures. Some of these relations will be determined by the inherent directionality features, either directly or by transitivity. The remaining relations will be determined by the LCA. Sheehan (2013b) expresses the role of LCA in the theory as follows:

\[
\text{(39) Revised LCA}
\]

(i) If a category A c-selects a category B, then A precedes/follows B at PF.

(ii) If no order is specified between A and B by the sum of all precedence pairs defined by (i), then A precedes B at PF if A asymmetrically c-commands B.
That is to say, the LCA functions as a last resort for linearization. This immediately captures the fact that while heads seem to be able to precede or follow their complements, specifiers always seem to precede the rest of the phrase. In this model, it thus follows from the fact that specifiers are never selected (see section 4.8 above). Another distinctive property of Sheehan’s theory is that a head and its projection form a single multi-segment category, distinct from the complement of the head. The technical device which Sheehan (2013b) proposes to achieve this effect is formalising projection of a categorial label as copying of the label. Consider the effect this has on the statement of linear relations, in, for example, (40), the structure of *have read the book*.

(40)

![Diagram](image)

Note that this is a graphic representation of structure only, not order. Here, because it is English, Aux is marked to precede the category it selects. This means it precedes the boxed category V. It does not entail that it also precedes D and N. V, in turn, is marked to precede the category it selects, which means it precedes the boxed category D, which does not entail that it also precedes N. However, in this case, because V precedes D, and D precedes N, V does precede N by transitivity. And likewise, because Aux precedes V, it will also precede D and N by transitivity, so the total linear order in this case is Aux>V>D>N, and no recourse to asymmetric c-command is necessary.

Consider now how this accounts for FOFC. First, consider the permitted disharmonic ‘inverse FOFC’ order in (41a), with the structure in (b).
(41) a. Kyllä Anne on auton ostanut.  [Finnish]
indeed Anne has car bought
‘Anne really has bought a car.’

b. \[
\begin{array}{c}
\text{Aux} \\
\text{Aux}_p \\
\text{V} \\
\text{O} \\
\text{V}_F \\
\end{array}
\]

(41b) is strictly a representation of structure. Aux is marked to precede the category it selects at PF, while V is marked to follow the category it selects at PF. This yields the following ordered pairs:

(42) \begin{align*}
O & > V \quad \text{(by selection)} \\
\text{Aux} & > V \quad \text{(by selection)} \\
\text{Aux} & > O \quad \text{(by the LCA)}
\end{align*}

total order: Aux>O>V

There is no selection-determined order between Aux and O, either directly or by transitivity. As such, the LCA steps in to order these two categories based on asymmetric c-command relations. Since Aux asymmetrically c-commands O, it will precede O at PF.

Now consider the ungrammatical, ‘FOFC-violating order’ in (43a), with the structure (43b).

(43) a. *Kyllä Anne ostan auton on.  [Finnish]
indeed Anne bought car has
The structure is the same as in (41) (it could be graphically represented with V to the left of O and Aux to the right of V). The only difference is the linear selection features on Aux and V: Aux is marked to follow the category it selects V, while V is marked to precede the category it selects O. The linear pairs are:

\[
\begin{align*}
V & > O \quad \text{(by selection)} \\
V & > \text{Aux} \quad \text{(by selection)} \\
\text{Aux} & > O \quad \text{(by the LCA)}
\end{align*}
\]

The total order: \( V > \text{Aux} > O \)

Note the total order resulting from this set of linear pairs in \( V > \text{Aux} > O \), rather than \( V > O > \text{Aux} \); since Aux asymmetrically c-commands O, it will precede it. The order seen in (43a) simply cannot be derived, given these selection features. This is, then, the explanation of FOFC in Sheehan (2011, 2013a,b). There is no narrow syntactic ban on ‘FOFC-violating structures’, it is just that these structures are never linearised with the complement intervening between the two heads in a head-final sequence.

One advantage of Sheehan’s system is that is can account for the fact that FOFC seems to hold both of base-generated and movement-derived structures. Because the head and its projections constitute a single category, it follows that where a complex specifier is involved, only the projecting head c-commands into the main clausal spine:
Thus in (45), V asymmetrically c-commands v, but O does not. Sheehan (2011, 2013b) argues that the only way to linearise FOFC-violating structures such as (45) is for O to be spelled out in its base position, yielding the order $V>v>O$. As there is fairly good evidence that movement of this kind is involved in some disharmonic languages, the fact that Sheehan’s approach can derive the same linearization effect here can be seen as a significant advantage. For an example of this, see Chapter 7 on the Head-Final Filter.

Another effect of this approach is that FOFC is predicted not to be sensitive to the complement/adjunct status of the material intervening between the two heads. Any XP which follows V in (45) or (43b) will pose the same linearization problem, whether it is a complement or an adverb. In chapter 6 we see evidence that this is an advantage of the approach as adjuncts also seem to behave like complements in this regard, so linking FOFC only to selection, as in Holmberg (2000a) is problematic.¹⁸

To explain the fact that DP complements of V are not subject to FOFC, whereas CPs (and arguably PPs – see chapter 5) are, Sheehan appeals to the fact that DPs are often strong islands. As FOFC is an effect only of linearization, the prediction is that atomised domains
will not display the effect. As extraction is very generally possible from embedded CPs and PP, but not DPs, this seems to make the correct cut without the need to appeal to extended projections or categorical distinctness. The strong prediction is that any surface FOFC violations will involve an atomised phrase in the following way (where outline font indicates atomisation):

(46)

\[ \text{Dp}_N \rightarrow V_F \]

By hypothesis, then, the reason why N can precede V in (46) is because it is not part of the same linearization domain. N has already been linearised with respect to D when the two are merged with V, meaning that they behave like an atom. While some challenges remain for this approach from the behaviour of preverbal PPs in Germanic languages (as in 34b), this approach seems promising as an explanation for certain kinds of FOFC-exceptions.

As Sheehan (2013b) shows, the need for externally merged specifiers to be linearised with respect to the main clausal spine also serves to provide a more nuanced PF-explanation of Huang’s Condition on Extraction Domain. Externally merged head-initial specifiers will always pose a linearization problem, requiring them to be immediately atomised and hence to behave like strong islands. Externally merged head-final specifiers, on the other hand, will not pose the same problems. Consider the following contrast:
In (47a), a harmonically head-final CP is externally merged in a specifier position. C is specified to follow T because of c-selection. Because the C-projection is an externally merged specifier, it is not c-selected and must precede the category v as it asymmetrically c-commands it. This gives the unambiguous linear order T>C>v by transitivity. Now compare (47b), where the externally merged specifier is harmonically head-initial. This time, C will be specified to precede T by c-selection. Again, the C-projection is not c-selected so its position with respect to v must be determined via asymmetric c-command. C again asymmetrically c-commands v and so must precede it. But these two orderings are not sufficient to order T with respect to v:

\[(48) \quad C > T, \ C > v\]

Moreover, here asymmetric c-command will not help, as T and v do not stand in any c-command relation. As dominance is not, in this theory, used in linearization (for principled reasons), the only option is for the externally merged specifier in (47b) to be atomised upon being externally merged. Sheehan (2013b) argues that this serves to explain the fact that many head-final languages fail to be subject to CED effects, something which other approaches to the CED have failed to explain.
A final prediction of this approach is that extraposition will be observed wherever a FOFC-structure is linearised. Rather than this being a FOFC-compliance strategy, it is a simple effect of linearisation. This provides a novel account of certain kinds of extraposition, which is empirically well-supported (see Sheehan 2011). A potential challenge is that extraposition does not always pattern exactly as expected. Preverbal head-initial CPs are extraposed wholesale rather than only the complement of C being extraposed, for example.

In Sheehan’s terms, one way to capture what is special about Finnish is that it has dual specification of V and Aux: \( V_P \) and \( Aux_P \) but also \( V_F \) and \( Aux_F \). The order \( V>Aux>O \) actually does occur in Finnish, albeit as a highly marked alternative: See Chapter 10 for a detailed discussion of Finnish. Note also the following consequence: If Aux in (43) is spelled out as a suffix on V (as is not uncommon for modal, aspectual and temporal functional heads), then the linearization of (43b) will have the effect of head-movement. This, too, will be discussed in Chapter 10.

4.11 FOFC as an effect of incorporation and pied-piping

A background assumption in the theory sketched in this section concerns the nature of head-movement or incorporation. It has some features in common with the theory in Holmberg (2000a) (see section 4.8) but differs from it in a number of crucial respects. Roberts (2010a) proposes that incorporation of \( Y \) to \( X \) where \( X \) asymmetrically c-commands \( Y \) is triggered just where \( X \) and \( Y \) Agree, in the standard way, and \( Y \)’s formal features are (properly) included in those of \( X \). Let us refer to this as *Agree-driven incorporation*. According to Roberts, this kind of incorporation underlies many cases of cliticisation, verb-movement, noun-incorporation, etc.

A second assumption, from Roberts (2010a), is the idea that the ‘core functional categories’ that make up the basic functional hierarchy in both clauses and nominals (and probably elsewhere, e.g. in the extended AP, although we will say nothing about that case here) contain progressively more
formal features as one moves up the hierarchy. So, the lexical verb, for example, has (at most) one formal feature, that which characterises it as a verb, namely V. The “light” v also has this feature, as well as φ-features. T also has a V-feature and φ-features, but in addition has its intrinsic Tense features. Finally, C has V-features, φ-features (which may be shared in one way or another with T, following the suggestion in Chomsky 2008), T-features (see numerous analyses of tense and sequence of tense stemming from Enç 1987) and its intrinsic clause-typing features (Holmberg 1986, Cheng 1991). So we have the following picture: 19

(49) Basic clause structure:

\[
[C[V, T, φ, Clause Type]] [T[V, T, φ]] [V[V, φ]] [V[V]]
\]

Analogously, we can think of the basic structure of nominals as follows (see Chapter 8):

(50) Basic nominal structure:

\[
[D[N, Q, φ, Def]] [Num[N, Q, φ]] [n[N, φ]] [N[N]]
\]

If we now take on Roberts’ proposals regarding incorporation, each head would have to incorporate with the “next head up”, which requires that the heads are in an Agree relation regarding these formal features. Let us suppose, however, that Agree in the standard sense is restricted to φ-features. Considering the clause structure in (49), this means that the only case where Agree-driven incorporation may take place is where T’s φ-set is included in that of C. On standard assumptions (following Chomsky 1995: ch. 4), v also has a set of φ-features, but this set relates to the object, and is therefore distinct from that of T/C.

At this point, we need to introduce a third assumption. Following Rizzi (2008), let us take c-selection to be external search, while Agree is internal search. C-selection, then, involves a head searching the numeration for another head bearing formal features of the relevant kind, 20 triggering external Merge with that head. We can now make a distinction between Agree-driven incorporation,
defined and discussed in Roberts (2010a), and select-driven incorporation, i.e. head-movement of a selected category to its selector. Given the assumptions above, incorporation through the functional hierarchy is always possible. In fact, according to Roberts (2010a) Agree-driven incorporation is obligatory. On the other hand, we postulate that select-driven incorporation is always in principle optional, subject to parametrisation.\(^{21}\)

The final postulate necessary for a new account of FOFC is as follows:

\[(51) \text{ Where incorporation of } X \text{ takes place, pied-piping of } XP \text{ is a parametric option.} \]

“Pied-piping” here refers, standardly, to the movement of the category containing the goal, in this case the goal of external search. Since the goal itself incorporates to the selecting head, the pied-piped XP must move to a position as close to that head as possible, internally merging with it to form a specifier.\(^{22}\)

Now consider the order of operations as an Extended Projection is built. After the VP is built, \(v\) c-selects the lexical head \(V\), and VP thus Merges with \(v\); this is a combination of External Search and External Merge. Next, we have the parametric option of incorporation of \(V\) to \(v\) (External Search and Internal Merge). Then, and only then, comes the option of pied-piping of \(YP\) to Spec\(X\) (Internal Merge). From this, we can see that the basic cases of FOFC now follow; the basic FOFC-violating structure in (52a) is underivable, since, wherever \(\alpha P\) moves to Spec\(\beta P\), \(\alpha\) must have incorporated into \(\beta\), as shown in (52b):

\[(52) \begin{align*}
\text{a. } & *[\beta P [\alpha P \alpha \gamma P ] \beta ] \\
\text{b. } & [\beta P [\alpha P (\alpha) \gamma P ] \alpha + \beta ]
\end{align*} \]

Consider once more the case of (S)AuxVO and its permutations. Aux-V-O can be derived as in (53a), with no incorporation, the order following from the LCA (see section 3.5). The order V-Aux-O (see Chapter 2, example (7), and Chapter 10, examples (4) and (5)) is derived as in (53b), by incorporation
of V into Aux with no pied-piping. The order O-V-Aux is derived as in (53c), by incorporation of V into Aux with pied-piping of VP.

(53) a. [Aux [V O]]
    b. [V+Aux [<V> O]]
    c. [AuxP [<V> O] [V+Aux <VP>]]
    d. *[AuxP [V O] Aux ]

The FOFC-violating structure/order (53d) cannot be derived if the system does not allow ‘comp-to-spec movement’ (i.e. movement of the complement of a head H to the specifier position of H) except in conjunction with incorporation, that is when realizing the option of pied-piping.23

In order to fully derive FOFC, more is needed. In particular, we have to require that incorporation can only “start at the bottom” of the extended projection, with the lexical head. However, there are clear cases in the literature of head-movement affecting only a relatively high position in an extended projection, e.g. v-to-T movement (possibly the correct analysis of English “have/be raising”; see Emonds (1976, 1978)), and T-to-C movement as in English subject-aux inversion; see also Chapter 11, where movement of higher heads in the clausal hierarchy is postulated without V-movement in analyses of Navajo and various other OV languages with tense/aspect prefixes (in 11.2.3).

Consider what would happen in a case of v-to-T movement associated with vP-pied-piping, but no V(P)-movement. The resulting structure for TP would be as follows:

(54)  [TP [IP (S) [v' (v) [vP V O ]]] [T v+T (vP) ]]

Assuming either v or T to contain an auxiliary, (S)VOAux order results, contra FOFC. Furthermore, T could incorporate to C with pied-piping, giving VO .. C.
As was discussed in section 4.9 on the theory of FOFC in BHR, given that select-driven head movement is dependent on categorial features defining extended projections it makes sense that it must ‘start at the bottom’, where the categorial feature defining the extended projection is introduced, however this notion is more precisely formalised (see section 4.9 and BHR for an attempt). Cases of head movement which start higher up are plausibly not cases of select-triggered movement. T-to-C in English, for example, involves polarity and interrogativity. In Holmberg (2003, 2016) it is analysed as Pol-to-C.

Note that this approach does not rely on a special movement-trigger for linearization. The parametric options for linearisation are:

\[(55)\]
\begin{align*}
&\text{a. merge} \triangleright \text{incorporate} \triangleright \text{pied-pipe} & \text{(head-final order)} \\
&\text{b. merge} \triangleright \text{(incorporate)} & \text{(head-initial order)}
\end{align*}

The operations merge, incorporate and pied-pipe are all intrinsically ordered in relation to one another. Note further that, if incorporation is intrinsically linked to affixation (see Tables 3 and 4), this also explains the greater incidence of affixation in OV language, since these languages must feature incorporation on the view developed here.

A further interesting consequence of this approach is that it provides a partial explanation of a generalisation discussed in Chapter 11, which is that head-final languages rarely have any prefixes. Consider the derivation in (56):

\[(56)\]
\[
[\cdot V X+V [XP (X) YP ] \rightarrow [VP [XP (X) YP ] [\cdot V X+V (XP ) ]]
\]

Here XP is the complement of V, and X incorporates to V, forming a prefix, followed by pied-piping of remnant XP, deriving a head-final VP with a prefix on V. But in terms of the above proposals concerning the categorial structure of extended projections, X cannot be part of the same extended projection as V; instead it is presumably in a high position in its own extended projection and therefore has either disjoint features in relation to V or a superset of V’s features. Either way,
incorporation of X is impossible under the present assumptions, and so (56) is not found. We do not, for example, find examples of D or C prefixed to V, with or without pied-piped NP or TP. This would be because D and C have more/disjoint features in relation to V and so do not incorporate.

We have now derived FOFC in a novel and interesting way, but we are faced with a different problem: how is complement-head order possible where incorporation is impossible? This problem arises from the fact that the theory allows XP-movement for linearization only where incorporation takes place. Incorporation is always an option as we build an extended projection, as we have seen, but is not possible “across” Extended-Projection boundaries, since either the relevant categories are disjoint or the lower head has more features than the higher one. For example, how can we derive OV order in a language like German, where it is clear that DP is head-initial and that D does not incorporate into V?

Here we have two options. The choice between them will be left open (it is possible, in fact, that different languages choose among them). The first option is to reinstate the Head Parameter, but in a more restricted form than the traditional formulations, e.g. in Hawkins (1983). Rather, along the lines of Richards (2004), it is relevant only for the relation between a lexical head and its complement; in fact, here a still narrower class of cases is relevant: the relation between a lexical head and a categorially-distinct complement; in all other cases word order is determined purely by movement. Thus, like Richards, we would adopt both the LCA and (a version of) the Head Parameter.

The second option is to modify the proposal above regarding the featural make-up of categories in the extended projection and allow, perhaps as a parametric option, that the “higher” positions in certain extended projections are either absent or have fewer formal features than we have been assuming up to now. Such a hypothesis seems implausible for a language like German, which clearly has initial C and D, along with final V, and so here the Head Parameter may have to be invoked. On the other hand, Japanese, for example, has no
clear examples of D and few really clear Cs (see Watanabe (2008) for some discussion of the Japanese DP), so perhaps the second option is at work in this language. More generally, given the relative paucity of clear D-final and C-final languages, it may be worth considering the possibility that the second option is found in the more harmonically head-final languages, with the Head Parameter at work only in disharmonic ones.

4.10 Conclusions

Word order is obviously one aspect of syntax where there is much cross-linguistic variation. The big question is, what are the limits of this variation, if there are any limits? What are the limiting factors? Is it UG?

This chapter has presented a brief history of the theory of linear order in generative Chomskyan linguistics, leading up to the currently widely accepted model where linear order is factored out of narrow syntax. In this model, narrow syntax builds hierarchies of sets of categories, by merge, deriving syntactic representations (LF-representations) which are semantically interpreted. Arguably these syntactic representations are universal, with no cross-linguistic variation possible. The hierarchies of sets of categories are assigned linear order as part of externalisation, the post-narrow-syntactic derivation of PF, subject to much cross-linguistic variation. As pointed out by Chomsky (1965), this model requires a theory of “how the abstract underlying unordered structures are converted into actual strings with surface structures” (Chomsky 1965: 125). Such a theory was provided in Kayne (1994), including the LCA, mapping asymmetric c-command relations to linear order. We have also briefly discussed alternatives to Kayne’s LCA-based theory, in particular Haider (2000), based on a universal branching direction combined with a notion of ‘licensing direction’, which is parametrized.
We then moved on to the structure-based condition on word order which is the topic of this book: FOFC. This condition, which, as we have seen in the preceding chapters, is pervasive across languages, does not follow from established notions of the relation between structure and order, and therefore presents a challenge for syntactic theory which has to be met. Four theories have been reviewed of the mapping of structure to word order purporting to explain FOFC, all compatible with the assumption that narrow syntax is order-free. One is the theory proposed in Holmberg (2000a), based on the classical version of the LCA according to which head-final order is necessarily derived by phrasal movement, combined with the idea that selection in the syntax requires/is mapped to adjacency at LF (the head-initial case) or PF (the head-final case). The second theory is articulated in BHR, also based on the version of the LCA according to which head-final order is necessarily derived by movement of the sister of the head, and is, in this sense, the marked alternative. In BHR’s theory, the mark of head-final order is a movement-triggering feature associated with the categorial feature [±V], defining extended projections. FOFC is an effect of how this movement feature ‘spreads’, or projects, throughout an extended projection via c-selection, subject to Relativized Minimality.

The third theory reviewed is articulated in Sheehan (2011, 2013a,b), and assumes that head-complement order is encoded as a lexically specified, inherent property of lexical categories. This lexical ‘directionality specification’ determines the order of a head and its complement at PF. However, because the grammar is universally required to unambiguously determine the linear relation between every pair of terminals in a syntactic domain, the lexical linear specifications are typically not sufficient. Asymmetric c-command is also required to order specifiers and disharmonic combinations of heads, explaining why it is precisely these things which display Kaynean Asymmetries. Sheehan assumes that the role of the LCA is to determine the linear relation between categories not ordered by virtue of these lexical
specifications. One consequence of this system is that it explains FOFC, in that the FOFC-violating order cannot be derived.

The fourth theory is based on the theory of head movement in Roberts (2010). As in Holmberg (2000a), the selection relation between a head $\alpha$ and its complement $\beta_P$, where $\alpha$ and $\beta$ are in the same extended projection, can be accompanied by head movement of $\beta$ to $\alpha$, or (unlike Holmberg 2000a) by head movement of $\beta$ to $\alpha$ with pied-piping of the remnant $\beta_P$ to $\text{spec,} \alpha_P$. If the second of these options is the only way a complement of $\alpha$ can end up as a specifier of $\alpha_P$, hence the only way it can end up immediately preceding $\alpha$, then FOFC will always be observed, because a FOFC-violating structure is underivable.

In subsequent chapters these theories, particularly the second and third one, will be put to use in the dissection of a variety of constructions exhibiting effects of FOFC, in a variety of languages. The four theories reviewed here are not the only ones in the literature purporting to explain FOFC. There are also proposals to the effect that FOFC can be explained in terms of conditions on speech processing; these are discussed in Chapter 5. Another recent proposal is articulated in Philip (2012) in terms of an Optimality Theory-based theory of syntax; see BHR for some critical remarks on this proposal.

---

1 Many thanks to Bob Freidin for his comments, which (I hope) helped to improve in particular the early part of the chapter. He is in no way responsible for the claims made here, though.

2 C-command can be defined in terms of dominance (or containment) and sisterhood: $\alpha$ c-commands $\beta$ if and only if $\beta$ is $\alpha$’s sister, or $\beta$ is dominated by $\gamma$ which is $\alpha$’s sister (Chomsky 2000), implying that the primitive structural relations are dominance/containment and sisterhood.
3 Typically complements and adjuncts occur on the same side of the head, although there are also many counterexamples (see Dryer 1992). This is consistent with the idea that there is one parameter for head and dependent, where complements and adjuncts are two types of dependents.

4 Chomsky (2000) argued that adjunction was a different type of merge, ‘pair-merge’, an inherently asymmetric process forming ordered pairs. While this idea has hardly been developed since, the notion that merge is an inherently asymmetric process always building ordered pairs has been articulated in work by Jan Wouter Zwart (see Zwart 2011a and the text below).

5 There is also lexical variation. Clearly, not all languages, have a word for, say, gooseberry (if they have never encountered or heard of gooseberries). More interestingly, not all languages have a word for religion (Boyer 2000), and not all languages have a definite article. A radical view is that all languages have a definite article, and possibly, in some sense, words for gooseberries and religion as well, but they happen not to have a pronounced form in some languages; see Ramchand & Svenonius (2008) for discussion (they do not take that position).

6 The notion of c-command assumed by Kayne (is (i) (Kayne 1994: 16):

(i) $X$ c-commands $Y$ iff $X$ and $Y$ are categories and $X$ excludes $Y$ and every category that dominates $X$ dominates $Y$.

The notion ‘category’ is crucial: while $X^o$ and $XP$ are categories, $X’$ level phrases count not as categories but as segments. This means that an $X’$ phrase does not c-command its sister $YP$, the specifier (or adjunct), while the sister $YP$ does c-command $X’$. This ensures asymmetric command between a specifier (or adjunct) and its sister, hence ensures that the specifier (or adjunct) necessarily precedes (the constituents of) its sister.

7 This is not to say that we concur with the view that labelling is required just to satisfy the LCA; labelling is crucial for the interpretation of any complex expression at LF as well (the semantic
difference between, for example, *wall paper* and *paper wall* is a difference in headedness, nothing else), and arguably is required for Merge to operate (that is if we reject Collins’s (2002) label-free syntax, and assume that c-selection requires labels). See also Sheehan (2013), Chomsky (2013).

See Sheehan (2013b) for a summary of further symmetry problems raised by BPS.

In bottom-up derivational terms, building a non-branching tree such as the NP in (6) requires an operation ‘grow’ or ‘sprout’, whereby N grows into an NP, but the only structure-building operation in BPS is merge.

Haider (2005, 2012) argues that there is a third type of language alongside VO and OV languages, namely, languages that are underspecified for licensing direction, and therefore allow VO and OV orders, employed in a systematic fashion.

Arguably, the Final-over-Final *Requirement* would have been a more transparent name – but FOFR is an even worse acronym than FOFC.

Svenonius (2006) suggests a way to reconcile these two hypotheses: The surface adjacency is not an output condition but a cue in the L1 acquisition of grammar (on ‘cues’, see Lighfoot 1999) leading the learner to a particular analysis, which in Svenonius’s terms is that there is an abstract head in the structure triggering movement. For example in (i) (which would be the derivation of (16c)) that head would be above the Aux, triggering movement of the complement of Aux.

(i) \[ F [_{A_{uxp}} \text{Aux} [_{VP} \text{O} [_{V} \text{V} <O>] ]]] \rightarrow [_{FP} \text{O} [_{V} \text{V} <O>] [_{F} [_{A_{uxp}} \text{Aux} <VP> ]]]

As argued by Abels (2003), if movement is triggered by the need to check a feature under locality, as in standard versions of minimalist syntactic theory, then movement from comp to spec cannot occur, since it moves a constituent from the most local relation possible (sisterhood) to a less local relation (spec-head). However, in BHR the movement in question is not triggered by the need to check a feature. The only effect it has, is changing the c-command relation between the head and its dependent, which in turn has head-final order as result.
14 A more neutral definition is ‘the sister YP of X° undergoes the minimal movement allowed by UG required to place XP in a position where YP asymmetrically c-commands X°’. This definition is neutral between (21a), assuming movement from comp,X° to spec,XP, and a theory of phrase structure where such a movement is ruled out as being ‘too local’ (see previous footnote).

15 The selection feature could be construed as an unvalued feature, being assigned a value by the complement (thus regarding categorial selection as a special case of Agree; Chomsky 2001). If so, the value assigned would not be the categorial value (since that is what is selected for), but the ‘value’ of the root. The identity of the root of L would in that way be carried along to the highest head in the extended projection of L.

16 As discussed by Sheehan (2013b) this is not an innocuous simplification: Including the subject means allowing multiple specifiers, excluded by the standard formulation of the LCA in Kayne (1994), relying on a category-based definition of c-command.

17 It is interesting to consider how a theory such as Haider (2000), briefly summarized above in section 4.6, where head-final order is basic and head-initial order invariably involves head-movement, can deal with FOFC. As mentioned, the head parameter is expressed in Haider (2000) as a difference in ‘licensing direction’: Heads license complements to the right or to the left. Licensing to the left is the less marked option, being harmonic with the universal branching direction. Left-licensing can proceed in principle without any movement: The head licenses its sister complement, the X’-projection of the head licenses its sister, the specifier or adjunct. If a head licenses to the right, on the other hand, it has to move to a position where it c-commands the complement (or specifier or adjunct). The right-licensing property of a head is thus effectively a head-movement-triggering feature. Let us symbolise it by \[ \rightarrow \]. The left-licensing property is unmarked. (i) represents the base-generated structure of the FOFC-violating configuration.
(i) \[[\text{AuxP} [\text{VP O V}]] \text{Aux} \]\[\rightarrow\]

The feature \([\rightarrow]\) will trigger head-movement of V, forming a VP-shell on top of VP, giving the typical FOFC-violating order (unless O moves even higher). In principle a version of BHR’s theory could be devised such that the marked \([\rightarrow]\) would be forced to ‘spread’ up the tree, to the highest terminal node in the extended projection (it would have to be forced to spread, not just allowed to, as in BHR). The unmarked left-licensing property, on the other hand, would not need to spread, which follows if it is the default, an effect of absence of a directional feature. The result would be the desired one: Wherever \([\rightarrow]\) is introduced in the tree, triggering head-movement and thus head-initial order, any higher node in the spine of the extended projection will also be head-initial.

18 Note that unlike earlier versions of the theory (Biberauer, Holmberg and Roberts 2007), the theory in BHR can accommodate the case of FOFC as an effect of adjunction. The L-movement-triggering feature triggers movement of the sister of L regardless whether it is a selected complement or a non-selected adjunct.

19 The notion of a “cartographic field” can now be defined as a sequence of functional heads with the same set of formal features; more formally, the functional sequence \(<F_1 \ldots F_n>\) such that all \(F_j\) have identical formal features. See Roberts (2010a) for the proposal that what distinguishes, say, the sequence of aspectual heads which Cinque (1999) argues constitute the lower part of the clausal functional sequence, are actually semantic features. Similarly, we can define the notion of Extended Projection as follows: \(\alpha\) and \(\beta\) form an Extended Projection where \(\alpha\) c-selects \(\beta\) and \(\alpha\) and \(\beta\) are non-distinct in lexical categorial features.

20 This can be stated, technically, by treating the selecting feature as uninterpretable, hence \(v\) has an uninterpretable V-feature, while V has an interpretable one, etc. (see Bazalgette (2011) on
this). More precisely, v has an uninterpretable Categorial feature, [Cat:__], which is valued in the standard way by copying V’s value for Cat. Interestingly, this appears to invert the more common view originating in Marantz (1997) that v “categorises” the acategorial lexical root. But it does not really, since we can still assume that roots are acategorial, and that each lexical head has the form [Cat Root Cat], quite visible in many languages although in English Cat is usually phonologically null. It may be that copying of this categorial feature that makes the fundamental difference between the two types of Extended Projection, entailing a different kind of external argument, different substantive features on the higher heads, etc. See also the Conclusion to Chapter Five.

21 Compare (17) in section 3.8. In that theory selection-driven incorporation is obligatory, but can be overt or covert.

22 This is a crucial difference between this theory and that in Holmberg (2000a) or BHR: There is no XP movement as such from complement position of a head H to the specifier position of H. What there is, is incorporation of the head of XP into H, with pied-piping of the rest of XP as an option subject to parametric variation, where the pied-piped phrase ends up as specifier of H.

23 The big advantage that this theory has over the one in Holmberg (2000a) in section 3.8 is that the dubious condition (20) is not required. Adjacency between the two heads in a head-final configuration is an automatic consequence of incorporation.