

## Chapter 5

### The Final-over-Final Condition and Processing<sup>1</sup>

Michelle Sheehan

#### 5.1 Introduction

It has often been claimed that word order universals lie beyond the remit of universal grammar and stem ultimately from parsing/processing principles or patterns of diachronic change rather than grammatical principles *per se* (cf. Hawkins 1994, 2004, Newmeyer 2005, Whitman 2008, Abels and Neeleman 2009, 2012). For this reason it is worth assessing whether The Final-over-Final Condition (FOFC), too, might stem from processing pressures operative either synchronically or diachronically. Note that the FOFC asymmetry raises some interesting issues in this regard as it is an asymmetry concerning the ordering of *heads*, whereas most other word order asymmetries concern the ordering of *specifiers/phrases* (cf. Cinque 2005a, Abels and Neeleman 2009, 2012, Richards 2008a). Abels and Neeleman (2009) claim that all the word order asymmetries taken as evidence for the LCA can be reduced to a ban on rightward movement, which has a direct motivation from parsing: the need for fillers to precede gaps. FOFC, notably, seems to be different in this regard as it is not immediately clear how it could derive from a ban on rightward movement. In this much, then, FOFC appears to be crucial evidence that the ordering of heads is also subject to the same kind of left-right asymmetry associated with the ordering of phrases. This is not to say, though, that FOFC requires the LCA. It is of course possible that some other processing pressure is directly (or indirectly) responsible for FOFC. At least two such potential explanations have been proposed in the literature: Hawkins' (1994) Early Immediate

Constituents and Cecchetto's (2013) restriction on backward dependencies. In this chapter we consider these two approaches and evaluate whether FOFC can be explained purely by processing pressures.

The remainder of the chapter is structured as follows. Section 2 introduces Hawkins' (1994) account of FOFC, which connects it to the more general preference for harmony in linguistic systems. Section 3 introduces a different approach developed by Cecchetto (2013), which attributes the effect to a restriction on backwards Agree, assimilating FOFC to Ross' (1967) Right Roof Constraint. Section 4 considers the predictions of these accounts and argues that they face certain serious challenges. Section 5 considers the way that processing interacts with grammar in the approaches in question and argues that even the accounts in chapter 4 might ultimately be attributed to a processing effect, albeit indirectly. Section 6 broadens the discussion to include phonology, identifying a striking similarity between the Onset Maximization Principle and FOFC. Finally, section 6 concludes and outlines some matters for future research.

## **5.2 Hawkins (1994)**

One highly articulated processing-based account of word order is Hawkins' (1994, 2004) Performance-Grammar Correspondence Hypothesis (PGCH). While Hawkins' approach is well known and highly influential as an account of the predominance of harmony in natural language, it is less well known as an account of the FOFC asymmetry (but see Cecchetto 2013, Sheehan 2013a for discussion). Interestingly, though, Hawkins' (1994) Early

Immediate Constituents, which gives rise to the preference for harmony, also provides an elegant account of (a statistical version of) FOFC.<sup>2</sup>

Hawkins (1994) proposes the following processing principle:

(1) Early Immediate Constituents (EIC) (Hawkins 1994: 77)

The human parser prefers linear orders that maximize the I[mmediate]C[onstituents]-to-non-I[mmediate]C[onstituents] ratios of Constituent Recognition Domains.

*Immediate constituents* (ICs) are the constituents required to identify a certain grammatical category/phrase. For example a transitive VP consists of two ICs: the category V, which is constructed by the verb and the category NP, constructed by the determiner or noun (assuming, as is standard in X-bar Theory that lexical items are distinct from category projections).<sup>3</sup> *Constituent Recognition Domain* and *IC-to-non-IC ratio* are defined in the following ways by Hawkins:

(2) Constituent Recognition Domain (CRD) (Hawkins 1994: 58-59)

The CRD for a phrasal mother node M consists of the set of terminal and non-terminal nodes that must be parsed in order to recognize M and all ICs of M, proceeding from the terminal node in the parse string that constructs the first IC on the left, to the terminal node that constructs the last IC on the right, and including all intervening terminal nodes and the non-terminal nodes they construct.

$$(3) \quad \text{IC-to-non-IC ratio} = \frac{\text{Number of ICs in domain}}{\text{Number of non-IC nodes in domain}}$$

As Hawkins acknowledges, the number of non-IC nodes in a given structure will vary depending on a number of independent syntactic assumptions (i.e. binary vs. ternary branching, heads and movements posited, and more pertinently nowadays X-bar theory vs. Bare Phrase Structure, vP-internal subject hypothesis). To avoid entering into such issues, Hawkins also provides a simplified IC-to-word ratio, which factors out syntactic assumptions and gives results broadly similar to the IC-to-non-IC ratio:<sup>4</sup>

$$(4) \quad \text{IC-to-word ratio} = \frac{\text{Number of ICs in domain}}{\text{Number of words in domain}}$$

Finally, he proposes that the ratio for a CRD is the average of the scores for all IC-to-word ratios calculated left-to-right.

(5) Calculating left-to-right IC-to-non-IC ratios

The L-to-R IC-to-non-IC ratio for a non-optimal CRD is measured by first counting the ICs in the domain from left to right (starting from 1), and then counting the non-ICs (or words alone) in the domain (again starting from 1). The first IC is then divided by the total number of non-ICs that it dominates (e.g. 1/2); the second IC is divided by the highest total for the non-ICs that it dominates (e.g. if this IC dominates the third through seventh non-IC in the domain, then 2/7 is the ratio for the second IC); and so

on for all subsequent ICs. The ratio for each IC is expressed as a percentage, and these percentages are then aggregated to achieve a score for the whole CRD.

(adapted from Hawkins 1994: 82)

According to Hawkins, these processing principles predict that harmonic head-initial and head-final constructions should be most common cross-linguistically as these kinds of structures are optimal in terms of processing, with CRD and ratios as small as possible. Consider, by way of illustration, the harmonic orders for VPs containing an adpositional complement and a verb (Hawkins 1994: 96-97):

- |     |   |   |                              |
|-----|---|---|------------------------------|
| (6) | [ <b>VP</b>   | [ <b>v</b> <b>go</b> [ <b>PP</b> [ <b>P</b> <b>to</b> ] [ <b>NP</b> school]]] | IC-to word ratio = 1/1, 2/2, |
|     | 1   | 2   | average= 100%                |
| (7) | [ <b>VP</b> [ <b>PP</b> [ <b>NP</b> school] [ <b>P</b> <b>to</b> ]] [ <b>v</b> <b>go</b> ]] | IC-to-word ratio = 1/1, 2/2   |                              |
|     | 1   | 2   | average =100%                |

In both (6) and (7) the NP complement of P is not included in the CRD of VP because NP is an IC of PP but not of VP. In (6), the IC-to-word ratio of V (the first IC) is 1/1, as the word *go* serves to construct it (though at this point it remains unclear whether the V is transitive/intransitive/ditransitive etc.). The IC-to-word ratio of PP is 2/2, as the second IC (PP) dominates only the second word contained in the CRD of VP, namely *to*. The average IC-to word ratio is therefore 100% as the number of words is exactly equal to the number of ICs constructed. In (7) a similar effect holds, except that this time, the head-final language is

constructed ‘bottom up’. Once again, the NP is not included in the CRD of VP. As such, the CRD begins with *to*, which constructs the first IC of VP, namely PP. Because NP is outside the CRD of VP, PP dominates only one word in the CRD (namely *to*), giving it an optimal 1/1 IC-to-word ratio. The word *go*, similarly, serves to construct the second IC of VP, and V also dominates the second word in the CRD (namely *go*). Once again the average of these two IC-to-word ratios yields a perfect 100% as two adjacent words serve to construct the two ICs of VP. As such, Hawkins’ approach means that harmonically head-final and harmonically head-initial structures are equally optimal in processing terms. If frequency correlates with processing efficiency, then Hawkins’ EIC theory predicts that harmonic structures will be most frequent in the world’s languages.

But EIC also yields more subtle predictions in relation to the two disharmonic combinations, as Hawkins (1994: 255) notes, where NP complements of P are necessarily included in the CRD of VP. He gives the following IC-to-word ratios, assuming that V and P are single words and that NP comprises a determiner and a noun:

(8) [VP [V go [PP [NP the shops] [P to]]]

1            2    3        4        IC-to-word-ratio = 1/1, 2/4 = 75%

(9) [VP [PP [P to] [NP the shops]] [V go ]]

1            2    3        4        IC-to-word ratio = **1/3**, 2/4 = 42%

In (8), the first word *go* serves to construct the first IC of VP, V, giving an IC-to-word ratio of 100%. The second IC of VP (PP) is constructed by *to*. Now PP dominates the second through fourth words in the CRD, and so the IC-to-word ratio of PP is 2/4. The aggregate IC-to-word

ratio is thus 75% for this word order. In (9) on the other hand, the first IC which is constructed is PP. In this case, PP dominates the first through third words in the CRD of VP (i.e. *to*, *the* and *shops*). According to the definition in (5), then, the first IC-to-word ratio in (9) is 1/3. The IC-to-word ratio of the second IC, namely V, is 2/4 because the IC V dominates the fourth word in the CRD (i.e. *go*). As such, (9) has a substantially lower efficiency rate of 42%. Moreover, the greater the number of words in the intervening NP constituent, the larger the difference in efficiency between the two disharmonic word orders will be. In effect, the NP complement of P is parsed twice in (9): once in the construction of the first IC (PP), and again in the construction of the second IC (V). Crucially, this makes the prediction that structures/orders like (9) will be more difficult to process, and hence less frequent than those in (8). As (9) is the FOFC-violating order, this version of the PGCH thus appears to derive a statistical FOFC from independently justifiable principles of efficient processing.<sup>5</sup>

In these terms, FOFC reduces to the fact that (i) CRDs are constructed left-to-right and (ii) higher heads are privileged in constructing more independent ICs. Another way to think of this is that, in a sequence head-phrase-head, the most economical way to parse it is head-[phrase-head] rather than [head-phrase]-head. The prediction, then, is that this parsing preference will be reflected in the linear orders of the world's languages. We will see below that these predictions seem to hold, though the numbers of FOFC violations are very small in all cases, arguably lower than predicted by Hawkins' account. However, the PGCH also makes other predictions, which are not so well supported. Crucially it predicts that where two categories display a typological preference for harmony they will also display a FOFC effect and vice versa.<sup>6</sup> This is because the same principle which gives rise to the preference for

harmony (i.e. EIC) also gives rise to FOFC. While there might be additional (historical/sociolinguistic) factors which skew the typological sample away from harmony, where this happens, these same factors should also serve to rule *in* a FOFC-violating order. The prediction of EIC is therefore biconditional:

- (10) Cross-linguistic preference for harmony between X and Y iff FOFC holds between X and Y

We will see below that there is suggestive evidence against such a biconditional relation. But first, we consider an alternative processing-based account of FOFC, proposed by Cecchetto 2013.

### 5.3 Cecchetto (2013)

Cecchetto (2013) derives the FOFC asymmetry from the following grammatical principle, which he takes to be operative in the syntactic component, presumably universally, motivated by processing pressures:

- (11) In a forward dependency, Agree can be cross-phrasal. In a backward dependency, Agree is phrase internal.

Assuming that Agree underlies all syntactic dependencies (cf. Chomsky 2000, 2001), it follows from (11) that any rightward dependency will have to be very local, whereas leftward



dependencies can be cross-phrasal, hence longer-distance.<sup>7</sup> This, Cecchetto claims, serves to assimilate FOFC to the Right Roof Constraint as both involve a backward Agree dependency, if, indeed, selection, like movement, involves Agree. In order for this to work it has to be that selection is a relation between heads in the normal case:

- (12) a. [YP<sub>i</sub> [ZP ... t<sub>i</sub> ]]  
 b. \*[[ZP ... t<sub>i</sub> ] YP<sub>i</sub>]<sup>8</sup>

- (13) a. [X [Y<sub>P</sub> Y ... ]]  
 b. \*[[Y<sub>P</sub> Y ... ] X]

The problem with FOFC, then, is that X in (13b) cannot backwards select a head in a lower phrase. In (13a), on the other hand, this is possible because the dependency is ‘forward’ and so can be cross-phrasal. Despite the initial appeal of this parallel, a number of questions arise in relation to this account. Firstly, there is a potential issue with the parallel itself. In filler gap dependencies the gap is dependent on the filler and not vice versa according to Cecchetto.<sup>9</sup> In selection, however, it is a head which, he argues, is dependent on its complement and not vice versa on the basis that truncation is only possible at the top of a given structure. Cecchetto provides the following licit example of truncation from Italian absolute small clauses, citing Belletti (1990), by way of example:

- (14) Mangiata        la mela,        Gianni morì  
          eaten        the apple       Gianni died  
          ‘Having eaten the apple, Gianni died.’

Although ellipsis of complements is of course also possible in many contexts, something which Cecchetto does not discuss, there is a sense in which the complement in ellipsis contexts like (15) is still present in the structure, unlike the truncated structure in (14):

- (15) John ate an apple and Mary did ~~{eat an apple}~~ too.

The elided VP is semantically present in (15) in that its content is recoverable from the antecedent VP in the preceding clause, whereas there is simply no tense information in the absolute clause in (14). There is a sense, then, in which the selection dependency can be argued to be unidirectional, like the filler-gap dependency, with a head being dependent on its complement and not vice versa. The particular parallel which Cecchetto draws in this instance, however, seems problematic. If filler>gap counts as a forwards dependency because the gap is dependent on the filler, as Cecchetto proposes, and gap>filler, conversely, counts as a backwards dependency, then comp>head should also count as forward dependency as it is the head which is dependent on its complement and not vice versa. This means that it should be head>comp which counts as a backwards dependency. But this seems to make exactly the wrong predictions, as in this case head>complement selection relations would be restricted to phrase-internal contexts and FOFC would be left unexplained.

Even leaving this potential objection to one side, moreover, there are further challenges for the approach. As Cecchetto (2013) acknowledges, one serious challenge is that (11), rather than ruling out [[head-complement]-head] orders *per se*, actually rules out selection between any two heads in a harmonic head-final sequence. In these terms then, there is nothing especially bad about FOFC; head-finality more generally is incompatible with straightforward head-head selection. How then does Cecchetto address the fact that harmonic head-finality is an extremely common property of natural language structures, arguably the most common word order? Cecchetto notes that this ban on selection only applies as long as selection is conceived of as a relation between two heads, as in (12b), adapted here:

$$(16) \quad *[[_{YP} Y_{[F]} \dots] X_{[uF]}]$$

According to (11), backwards Agree must be phrase-internal and X is external to the phrase containing Y. Cecchetto claims, however, that in such cases a selection relation *is* possible between X and the entire phrasal complement YP. In such contexts, crucially, head-movement, the usual means of creating fusional categories, is predicted to be banned as there is no head-head dependency. As such, it is predicted that head-final orders will be incompatible with head-movement and/or fusional morphology. As evidence for this prediction, Cecchetto notes a statistical correlation between head-finality and agglutinating morphology.

A potential problem with this solution is that it appears to undermine this approach as an account of FOFC as in such cases the selected phrase could be either head-final or head-initial, in the latter case in violation of FOFC:

- (17) a.  $[[_{YP [F]} ZP Y] X_{[uF]}]$   
 b.  $[[_{YP [F]} Y ZP] X_{[uF]}]$

As Cecchetto notes, the possibility of phrasal-selection in such contexts appears to provide an account of the fact that FOFC fails to hold across extended projections, if one makes the assumptions that verbs select DPs, rather than Ds. If phrasal selection is generally available in harmonically head-final systems, though, why is it not always available in the FOFC-violating structure? What prevents an auxiliary verb from selecting VP, rather than V, ruling in  $*[[V-O]-Aux]$  structures? This problem seems particularly serious in systems displaying variable word order, such as Old English, Hungarian, Basque and Finnish, where head-final sequences are perfectly possible and yet the FOFC-violating order is nonetheless banned.

## 5.4 A comparison of the various approaches to FOFC

### 5.4.1 On statistical FOFC effects

Both Hawkins and Cecchetto propose that the FOFC-violating order is ruled out because it is difficult to parse. For Hawkins this is because it is an inefficient way to construct immediate constituents online: a phrase intervening between two heads is better parsed with the head that follows rather than that which precedes it. For Cecchetto, the selection dependency, like all dependencies, must operate left to right as this is the order of the march of time, and only very local dependencies are possible in the opposite direction. For Cecchetto, though, FOFC arises because of a hard-wired principle of UG and so the effect is predicted to be absolute.

For Hawkins, on the other hand, the FOFC order will be dispreferred by the parser but not ruled out by UG. As such, the PGCH, unlike Cecchetto's approach and those in chapter 4, does not completely rule out the possibility of FOFC-violating orders. Rather it predicts that they will be infrequent and certainly less frequent than harmonic or inverse-FOFC orders. Presumably, over time, such orders are predicted to die out, under the pressure to be reanalysed as more efficient word orders.

Moreover, IC-to-word ratios are also influenced by the relative 'heaviness' of the constituents involved and the implication is that average relative weights might affect grammaticalization trends (cf. Hawkins, 2013, for discussion). As such, the PGCH predicts that whether the structure [[Y-ZP]-X] is permitted may be (i) directly sensitive to heaviness of ZP or (ii) sensitive to the category of Z, where different categories have different average weights. For example, the FOFC-violating order would be expected to surface more frequently with lighter categories like DP and less frequently with heavier categories like CP (which typically contain more constituents/words).

#### **5.4.2 Exceptions to FOFC**

It remains to be seen then, whether EIC can successfully account for apparent exceptions to FOFC. In their discussion of FOFC, Biberauer, Holmberg & Roberts (2014) note certain kinds of counterexamples and propose that FOFC holds only within extended projections. The reason they introduce this proviso is that DP complements of V, which surface in a preverbal position, are head-initial in some well-studied familiar languages like German:

- (18) Johann hat                    das/ein Buch   ausgeliehen  
        John    has                    the/a book     borrowed  
        ‘John has borrowed the/a book.’

In actual fact, there are relatively few OV languages with clear determiners in which DPs surface as the complements of V. Most OV languages either (i) lack determiners distinct from demonstratives (which are plausibly specifiers rather than heads) or (ii) have final determiners (cf. Dryer 1992:104). It is also noteworthy that it has been claimed that all OV languages permit scrambling (Kayne 2004, though the Ugric languages Khanti and Mansi seem to be rare counterexamples to this trend). Under one approach, scrambling, like differential object marking more generally, targets all DPs, with only structurally truncated NPs remaining in situ (Richards 2008b). If this analysis is along the right lines then D-initial OV languages with initial the relevant DPs surfacing inside VP may not actually exist, making a more general version of FOFC possible. This raises the question of the status of indefinite articles. Sheehan (2011) argues that they are specifiers rather than heads, so that the indefinite in (18) may also fail to counterexemplify FOFC. If this is the case then it would explain the fact that OV languages with initial articles are no less common (taking into account areal and genetic factors) than VO languages with final articles, without any need to add stipulations to the FOFC generalisation.

Of course Hawkins’ approach is concerned primarily with surface word order rather than syntactic structure so the fact that DPs often scramble in OV languages will only be relevant to the extent that additional material between a verb and its complement creates longer CRDs. Hawkins rejects Abney’s (1987) DP-hypothesis and assumes that NPs

comprise Det and N in a flat structure, with either being sufficient to construct an NP. As such, N-initial and Det-initial nominals are equally efficient in processing terms, as both can effectively function as the head of NP.<sup>10</sup> As Hawkins (2013) notes, though, initial determiners will carry a processing advantage only in non N-initial VO languages; in N-initial languages, they will make no difference and in non N-initial OV languages they will actually increase the CRD of VP. As such, the order D-NP-V *should* arguably be dispreferred compared to V-NP-D, as the former but not the latter carries a processing disadvantage. It seems, then, that EIC *does* predict a FOFC effect in this domain, despite the flat structure assumed for NP/DP. Recall, however, that EIC is sensitive to the relative heaviness of phrases. It follows, therefore, that DPs, tending to be light, might allow FOFC-violations more often than other heavier categories like PP and CP from a cross-linguistic perspective.<sup>11</sup> The lack of a skewing between the two disharmonic combinations of article, NP and V might not, therefore, be a serious problem for EIC.

A complication for this view arises, however, from Dryer's (1992: 103) observation that articles are nonetheless verb patterners. It is not clear why the lightness of a given phrase would be sufficient to render it immune to FOFC but not to the preference for harmony. Of course, the preference for harmony in this domain might have some independent (diachronic) motivation, but it is implicit in Hawkins' approach that diachronic processes will also be motivated by EIC. The facts seems to suggest, then, that FOFC and the preference for harmony might not stem from the same processing principle, contrary to the predictions of EIC. We return to this issue below.

Biberauer, Holmberg & Roberts (2007, 2008, 2014) also discuss another kind of exception to FOFC involving prepositional phrases in OV languages. Once again, German provides a relevant example:

- (19) Sie ist [<sub>VP</sub> [<sub>PP</sub> nach Berlin] gefahren]  
she is to Berlin driven  
'She went to Berlin.'

Despite a small number of robust counterexamples, though, it appears that there is cross-linguistic evidence of a FOFC skewing between PP complements and V. Firstly, the raw data from Dryer (2015a,c) indicate that there are more inverse FOFC than FOFC-violating languages, and that these languages occur in more macro-areas and represent a more genetically diverse group than the small number of FOFC-violating languages:

[Table 4.1 here]

These data seem to be consistent with the spirit of EIC, whereby FOFC is a statistical rather than a categorical effect, as there appear to be a number of languages which allow the FOFC-violating combination, though this order is dispreferred. Note, however, that *WALS* contains data about the directionality of PP and VP but it does not consider the actual placement of head-initial and head-final PPs in otherwise OV and VO languages. An examination of the P-NP OV languages reveals that many of the 12 languages, despite being OV, actually require



PPs to be obligatorily postverbal, meaning that the surface FOFC-violating construction \*[[P-DP]-V] does not actually occur.

Of the 12 languages, only the Indo-European languages (German, Dutch, Persian, Tajik, Kurdish and Sorbian) and the Semitic language Tigré appear to allow the word order P-DP-V. The other languages all either lack true adpositions or are languages in which PPs appear in a post-verbal position. Mangarrayi (Australian, Mangarrayi, Northern Territories) lacks true prepositions according to Merlan (1989: 26), instead it has ‘prepositional-like phrases consisting of an adverb followed by a noun appropriately case-marked to complement the combined meaning of the adverb and verb in the clause’. Moreover, the order of the adverb and its complement is not fixed; both *adv>comp* and *comp>adv* are possible. As such it is not clear that Mangarrayi is a true exception to FOFC. As Sheehan (2013a) notes, moreover, in Pāri (Nilo-Saharan, Nilotic, Sudan), Tobelo (West Papuan, North Halmaheran, Indonesia), Iraqw (Afro-Asiatic, Southern Cushitic, Tanzania) and Neo-Aramaic (Afro-Asiatic, Semitic, Israel) PP complements to V seem to surface in a post-verbal position, as the following examples illustrate:

- (20) á-lw’<sub>ΛΛr</sub>      kí      kwàc      [Pāri]  
          1s-fear      PREP leopard  
          ‘I am afraid of leopards.’      (Anderson 1988: 303)

- (21) lābulmunne      [ta-Bagdàd]      [Neo-Aramaic]  
          Take.me      to-Bagdad  
          ‘Take me to Bagdad!’      (Khan 1999: 338)

- (22) i-na      ta'<a'>ín                  [ay dí-r                  konkomo]    [Iraqw]  
S-PAST run<HAB>3SGM    to place.F-CON cock  
'He ran to the cock'                                  (Mous 1993: 100)

In Pāri, this is obligatory (Anderson 1988: 303), as it is in all dialects of Neo-Aramaic (Geoffrey Khan, personal communication). There is insufficient information to say the same for Iraqw, though all examples suggest it to be the case.<sup>12</sup> Matters are slightly less clear in Tobelo.<sup>13</sup> This means that the number of languages displaying FOFC-violations of this type is extremely small so that [[P-DP]-V] is far rarer than the inverse-FOFC order [V-[DP-P]]. If this were not a FOFC effect, but merely an effect of the preference for harmony then there would be no explanation for the fact that postpositional phrases in VO languages are rarely preposed (cf. Dryer 1992: 92).

As such, in a sense, a statistical FOFC-asymmetry appears to be attested between PP and V and this is potentially evidence against approaches which predict FOFC to be absolute and in favour of EIC as an analysis of FOFC. That said, the fact that the counterexamples are

so small in number and largely genetically related means that an independent explanation for their exceptional behaviour may well be possible, and the condition need not yet be abandoned as a hard condition. Whether this is plausible more generally depends on the status of counterexamples in other domains.

Bazalgette, Roberts, Sheehan and van der Wal (2012) take all the head-complement pairs from *WALS* and compare them pairwise, relative to their hierarchical rankings. As FOFC holds transitively through the clause, any head-initial phrase which is lower than a head-final phrase represents a FOFC-violation. Inverse FOFC is the reverse disharmonic combination and the two harmonic combinations speak for themselves. Figure 1 shows the results of this comparison irrespective of extended projections:<sup>14</sup>

[Figure 4.1 here]

What these data seem to suggest is that there is a very general FOFC effect, with the exception of TA/PP and case/number combinations. In all other cases, the harmonic head-final combination is dominant and the inverse FOFC order is strongly preferred to the FOFC order. In most cases the number of counterexamples is very small indeed, but again, they require examination if one wants to maintain the idea that FOFC is a hard constraint. Note also that in some cases the inverse FOFC order is preferred over the harmonic head initial order. As such, these data present a different problem for Hawkins's approach: there appear to be instances where FOFC holds between X and Y without X and Y displaying any tendency for harmony. This is problematic if both stem from the same processing principle, something which we turn to in the following section.

### 5.4.3 The relation to harmony

According to the PGCH, harmony and FOFC are two effects of the same processing principle (EIC). The clear prediction, given in (10), is that if two categories show a preference for harmony then they should also show a FOFC skewing and vice versa. The other approaches to FOFC, including Ceccehtto's approach, makes no such prediction. Rather they predict that the preference for harmony between two categories may be neutralized for some independent reason, but FOFC will always hold.

Some categories display both a preference for harmony and a FOFC effect. This is the case for Aux and VP, V and PP, C and VP and V and CP (Dryer 1992). As such, these pairs of categories are well behaved by the standards of EIC. Potentially more problematic for EIC is the fact that according to Dryer (1992) articles are also verb patterners, but not subject to FOFC effects cross-linguistically (see previous discussion). Polar question particles seem to display the opposite pattern. From a functional perspective, question particles (Pol) arguably serve to identify a polar question, and C serves to construct an embedded clause. For this reason, it is surprising that according to the data in table 1 from Dryer (2015d,e) repeated here, the disharmonic inverse-FOFC combination initial C and final Pol is more common than the harmonic combinations in terms of languages and genera and almost as common in terms of language families.

[Insert table 4.2 here]

There is thus no evidence of a preference for harmony between C and Pol. Crucially, though, FOFC appears to hold in this domain (see chapter 2). Once again, then, we witness the independence of harmony and FOFC: the same two categories displays a FOFC effect but no preference for harmony. This is highly problematic for EIC, which attributes both FOFC and the preference for harmony to the same processing principle. Of course, it remains an open question why there is no harmonic preference between Pol and C.<sup>15</sup> What is clear though is that in such cases, FOFC effects cannot straightforwardly be attributed to EIC.

#### **5.4.5 Interim conclusions**

As such, it is arguably the case that Cecchetto's processing-based account of FOFC fares better empirically than Hawkins', as it does not tie the effect to a preference for harmony. On a deeper level, though, Cecchetto's approach faces many challenges. Firstly, Hawkins captures the FOFC asymmetry without attaching any specific parsing cost to harmonically head-final systems, whereas Cecchetto is forced to assume a wholly distinct selection mechanism in head-final systems. While Cecchetto's observation that many head-final systems are agglutinating or lacking in head-movement is interesting, the idea that head-finality might pose a parsing difficulty seems unmotivated. The availability of this selection mechanism is thus suspect and actually serves to undermine the approach as an account of FOFC. Secondly, Hawkins' attempt to assimilate the FOFC effect to the preference for harmony, while empirically problematic for the reasons discussed above, is still less problematic than Cecchetto's attempt to assimilate FOFC to the Right Roof Constraint. As discussed above, the right roof constraint seems to be opposite to head-finality (and FOFC) in

some respects as it is a case where a filler cannot follow a (dependent) gap, whereas head-finality is a case where a head (dependent) cannot follow its complement. As such, both of the processing-based accounts of FOFC that we have discussed face certain serious problems, though this does not mean that no processing-based account of the effect is possible. Indeed, the fact that a statistical version of FOFC appears to hold across extended projections and even, as we shall see below, at the level of phonology, strongly suggests that the ultimate explanation for FOFC may well stem from parsing. We discuss this matter further in the following section, where it will further be proposed that the existing ‘formal’ accounts of FOFC may be attributed at some level to the preferences of the parser.

## **5.5 The effect of the parser on grammar**

Another important difference between the two approaches to FOFC discussed here is the relationship they posit between the parser and the narrow syntax. For Cecchetto, the restriction on backwards dependencies is a (universal?) principle of grammar which ‘biologises’ a preference of the parser (to use Kiparsky’s 2008 term). For Hawkins, the relationship between the parser and the grammar (of a specific language) is actually more indirect and the claim appears to be that languages will change over time according to the preferences of the parser, so that this relationship is always mediated by language use. As such the innate component of grammar itself is not affected by parsing pressures, only the grammars of specific languages are (cf. also Newmeyer 2005, 2006). Of the two possibilities, Hawkins’ seems the least problematic. The idea that principles of universal grammar could be functionally motivated seems problematic and inconsistent with the apparent fact that

language evolved very quickly, without intermediate stages of selection of the kind proposed by Pinker and Bloom (1990). But Hawkins' approach is not all that satisfactory either as if the relationship between the parser and the grammar of particular languages were always messily mediated by 'language use' then we might actually expect to find more counterexamples to FOFC than we actually do. More generally, we might expect asymmetries of the kind discussed by Kayne (see Chapter 4, section 4) to be less striking and more statistical than absolute. Of course, it is true that there are some apparent counterexamples to FOFC discussed in Chapters 4, 9 and above, but these seem to be fairly small in number and the condition seems to hold in several instances where languages have otherwise very free word order.

Biberauer, Roberts & Sheehan (2014) and Biberauer, Holmberg, Roberts & Sheehan (2010) propose a role for the parser which falls between these two extremes, whereby parsing preferences shape the grammar of a particular language within the limits imposed by UG by exerting a pressure during acquisition. This has the effect of creating a *No-choice parameter* whereby the language learner is given 'an offer she cannot refuse'. They begin with the observation that formal accounts of word order asymmetries which rely on the LCA face the problem that the form of the LCA itself seems to be a stipulation. Is it a principle of UG? It is easy to see why natural language would have to rely on syntactic relations such as c-command in the mapping from hierarchical structure to linear order. What does not follow from anything is why asymmetric c-command should map to precedence rather than subsequence. It is also easy to see that the fact that asymmetric c-command maps to precedence rather than subsequence (the two logically given options) might stem from parsing pressures (see Crain and Fodor 1985). However making the LCA a principle of UG in

this way effectively biologises the requirements of the parser, as in Cecchetto's approach. Biberauer, Roberts & Sheehan (2014) and Biberauer, Holmberg, Roberts & Sheehan (2010) propose, as an alternative, that the LCA (or the revised LCA in Sheehan 2013a,b) is an emergent parameter of UG. Every language must get from hierarchical structure to linear order using syntactic relations. Whether these syntactic dependencies map to precedence or subsequence is simply not specified by UG and the child must decide which option to take during acquisition, as a parametric choice. The parser, which exerts a pressure on language acquisition, strongly prefers the precedence option and so the LCA is effectively a no-choice parameter. As such, it could be argued that the preferences of the parser are the ultimate explanation for FOFC (and other word order asymmetries). The parser forces a linearization parameter to be set to precedence for parsing reasons, triggering a plethora of more arbitrary word-order asymmetries as side-effects, notably FOFC.

## **5.6 The Onset Maximization Principle<sup>16</sup>**

Thus far, it has been argued that the two existing processing-based accounts of FOFC face certain problems but that 'formal' accounts based on the LCA (discussed in Chapter 4) might actually be ultimately grounded in the requirements of the parser. An interesting parallel to FOFC found in the domain of phonology suggests that matters may be even more complex than this and that there may be more to be said about the ultimate explanation for FOFC, an issue which we leave open here.

If we assume that vowels/nuclei are equivalent to syntactic heads and consonants are equivalent to phrases then a very well-known proposed principle of phonology turns out to be



very similar indeed to a linear version of FOFC. The Onset Maximization Principle prefers consonants to be grouped with a following rather than preceding nucleus (as an onset rather than a coda, subject to partially language-specific prosodic/phonotactic constraints) (Blevins 1995: 230):

(23) *\*/...VC-V.../ /...V-CV.../*

The preference in (23) for consonants to form a constituent with a following vowel in such contexts seems analogous to the preference to parse a phrase with a following head rather than one preceding it. As Itô (1989) shows, things get more complex (and language-specific) where an intervening consonant cluster is involved, but where a single intervening consonant is present, the effect is fairly robust, a fact that is reflected typologically by the fact that many languages require all syllables to have onsets but no known languages requires all syllables to have codas (Itô 1989: 222).<sup>17</sup> As such the sequence VCCV can be syllabified as /V-CCV/ or /VC-CV/, but not \*VCC-V, as he notes (*ibid.*).

While Onset Maximization seems tantalizingly similar to the FOFC asymmetry, it is not clear whether a unified account of both effects is possible or desirable. This is true, in part, because of the many obvious differences between syntax and phonology. One obvious difference is that syntax, unlike phonology is hierarchically structured and recursive (see Tallerman 2006 for discussion). Although the syllable has a constituent structure, the only sense in which syllables might be considered to be hierarchically structured in relation to each other is in terms of prosodic features such as stress. The parallel is thus not perfect, as there seems to be no requirement for the final V in (23) to dominate the initial V in any real sense.

If a unified account is to be provided for Onset Maximization and FOFC, then, it would need to be a linear account and it is not clear whether a linear account will work for the syntactic version of FOFC, for the reasons outlined above and in Chapter 3.

It might be insightful, nonetheless, to consider previous accounts of Onset Maximization and examine whether they can be extended as analyses of FOFC. Unfortunately, there are not many such accounts as Onset Maximization has largely been taken to be directly encoded in UG, either as, for example, in Steriade's (1982) 'Universal Core Syllable Rule, which required all syllables to be of the type CV or in a more nuanced parameter, Itô's (1989) Onset Principle which states that languages either *avoid* or *ban* onsetless syllables. Even in more recent Optimality Theoretic approaches, Onset Maximization is simply encoded as a universal preference (Prince and Smolensky 1993). Of course, it might be that Onset Maximisation stems from some deeper parsing principle, plausibly connected to salience, but we leave a full investigation of this possibility to future research.

## **5.7 Conclusions and unresolved issues**

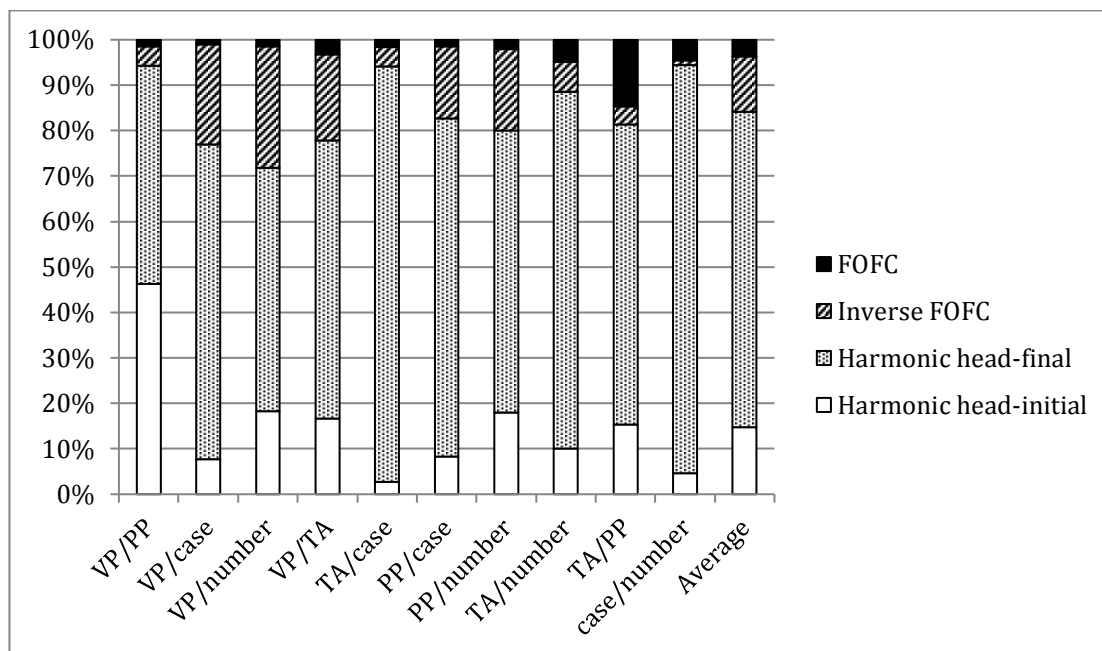
In this chapter, we have considered two potential accounts of the FOFC asymmetry stemming from processing pressures. It has been argued that while both have certain (different) merits, they nonetheless face certain (different) problems. Hawkins' account, based on EIC, faces the problem that there does not appear to be a biconditional relationship between the preference for harmony between two categories and a FOFC-skewing between those two categories. Cecchetto's approach does not face this empirical problem, but faces certain other challenges,

notably from the frequency of head-final orders and the existence of FOFC-effects even in languages otherwise permitting head-final sequences. It has nonetheless been argued that the ultimate explanation for FOFC might stem from parsing pressures, if some version of the LCA can be rethought as a no-choice parameter along the lines outlined by Biberauer, Roberts and Sheehan (2014).

Some issues, nonetheless remain unresolved. The Onset Maximization Principle, it has been noted, bears certain superficial similarities to FOFC and this coupled with the fact that FOFC effects can be observed between pairs of categories in distinct extended projections, suggests that a syntactic approach to FOFC may not be general enough. It is hoped that future research can address these questions and shed new light on the differences and similarities between phonology, morphology and syntax.

**Table 4. 1: Languages with disharmony between the ordering of VP and PP**

	FOFC-violating P-NP and OV	Inverse-FOFC NP-P and VO
N of language	12 <sup>18</sup>	38
N of macro areas	3	5
Language families	5	16
Genera	8	22



**Figure 4.1:** the four word order combinations across categories (from Bazalgette, Roberts, Sheehan and van der Wal 2012)

**Table 4.2:** Typological positioning of Polarity heads and complementizers

Type	Position of Pol	Position of C	Number of Languages (genera: families)
A	Initial	Initial	72: 35: 13 (78)
B	Final	Final	45: 33: 20 (46)
C	Final	Initial	74: 40: 16 (82)
D	Initial	Final	4: 3: 3 (4)

---

<sup>1</sup> This chapter is partly based on Sheehan (2013a), a previous version of which was published in Newcastle Working Papers in Linguistics. We thank the reviewers of both those publications for forcing us to clarify the presentation in various respects.

<sup>2</sup> The fact that Hawkins (2013) rejects FOFC as an empirical generalization is, thus, in one sense, surprising, and marks a change between Hawkins (1994) and Hawkins (2004). In personal communication, Hawkins notes that his 2004 approach could explain the asymmetry only if there were some preference for mother attachment over daughter attachment. Interestingly, this is reminiscent of the core idea behind the Linear Correspondence Axiom whereby there is a requirement for higher categories to precede lower ones. See Walkden (2009) for further discussion of Hawkins (2004) in relation to FOFC.

<sup>3</sup> Hawkins rejects the DP-hypothesis and uses NPs and we follow his usage here for consistency.

<sup>4</sup> As far as we can see, Hawkins gives no definition of what a word is, and rather uses the term in an informal sense. This may, of course prove problematic, especially if Julien (2002) is right, and words have no real syntactic status. See also chapter 12 for discussion of FOFC effects below the word level, which supports Julien's general approach. Recall, though, that the technical statement of his principle is in terms of IC-to-non-ICs, so this problem may not be all that serious. One might also plausibly reformulate his proposal, moreover, substituting 'morpheme' for 'word'. As we shall see in section 5, this may well be necessary on empirical grounds.

<sup>5</sup> DP and PP complements to V are beyond the remit of FOFC for Biberauer, Holmberg & Roberts (2008, 2014) because of a category or extended projection proviso. Nonetheless, as

---

discussed below, there is a sense in which a kind of FOFC-asymmetry is observed with PP complements of V, suggesting that PPs should perhaps not be ruled out of FOFC. Sheehan (2011) also discusses a potential FOFC-effect between V and truncated nominal NumP complements, resulting in extraposition. See also Chapter 4.

<sup>6</sup> We are not concerned here with conceptual objections to Hawkins' general approach, but rather with teasing out its empirical predictions. For a conceptual critique of the PGCH cf. Mobbs (2008).

<sup>7</sup> See Carstens (2013) for an implementation of a very similar idea.

<sup>8</sup> It is not totally clear how Agree can mediate extraposition. If the assumption is that all A-bar movements are licensed by Agree and extraposition is an A-bar movement, then the problem remains that the head 'attracting' the extraposed phrase (v in (i)) can be seen to be initial in languages like English, and yet the Right Roof Constraint holds nonetheless:

- (i) I [<sub>VP</sub> met a man t yesterday] [who I used to know]

Even if the specifier of v is to the right, the Agree dependency is presumably between v and the trace and is hence a forward dependency.

<sup>9</sup> As such there are conceptual as well as technical issues with reducing extraposition to an Agree relation as the dependency goes in the wrong direction: a trace is dependent on its filler, not vice versa.

<sup>10</sup> Hawkins uses the label Art rather than Det.

<sup>11</sup> Heaviness equates to number of words for the purposes of online reordering processes. According to the PGCH, though, grammars can develop categorical rules stemming from heaviness tendencies. Thus the fact that PP always contains DP makes PP heavier than DP in

---

general terms, even though both can have different heaviness values in different contexts. As such, an extraposition rule might be grammaticalised to apply only to PP and not DP irrespective of their relative online heaviness.

<sup>12</sup> Following Kandybowicz and Baker's (2003) analysis of Nupe, we might take the position of PP as evidence that these languages are basically head-initial with DP-V order derived via A-movement.

<sup>13</sup> The vast majority of the examples show PPs to be postverbal:

- (i)      ngohi-o                      to-modeke      de      o-Matias  
             I-also                      1-agree              PREP      NM-Matias  
             'I also agreed with Matias.' [Tobelo, Holton (2003:30)]

However, Holton (2003:55) explicitly states that oblique arguments can occur 'either before or after the verb'. Unfortunately, he uses the term 'oblique arguments' to refer to both PPs and DPs marked with a locative or directional case suffix. The only example with PP-V order which he gives involves an instrumental PP which may be a topic in the CP layer:

- (ii)      de      ma-kakatama      n-a-lye-ino      [Tobelo, Holton (2003:55)]  
             P      NM-tongs      2-3-roll-ALL  
             'Roll it up with the tongs.'

In the absence of further data is it impossible to say for certain whether Tobelo PPs adhere to FOFC or not.

<sup>14</sup> TA = tense/aspect marker

<sup>15</sup> Ultimately this might also shed light on the reason why final Q is so frequent in VO languages (cf. Dryer 2015a,e).

---

<sup>16</sup> Many thanks to Bert Vaux and Francis Nolan for useful discussion of the issues in this short section. They should, of course, not be taken to endorse any of what follows.

<sup>17</sup> Francis Nolan (p.c.) notes that there are empirical problems with Onset Maximization in its strongest form, even in English.

<sup>18</sup> We have added German and Dutch to this category although they are categorized as having mixed order of verb and object on WALS because of their V2 property.