SQUIBS

Deriving the Final-over-Final Constraint from third factor considerations[*]

George Walkden

Department of Linguistics, University of Cambridge

In this squib I examine two superficially competing explanations for the Final-over-Final Constraint (FOFC): those of Hawkins (2004) and of Biberauer, Holmberg & Roberts (2007, 2008). I argue that while an external, quantitative approach cannot account for all the relevant facts, such an approach, correctly formulated, may play a role in explaining the origin of a grammar-internal principle.

In this squib I examine two superficially competing explanations for the Final-over-Final Constraint (FOFC): those of Hawkins (2004) and of Biberauer, Holmberg & Roberts (2007, 2008). I argue that while an external, quantitative approach cannot account for all the relevant facts, such an approach, correctly formulated, may play a role in explaining the origin of a grammar-internal principle.

Section 1 discusses the issue of explanation in syntax. In 2 I give an overview of FOFC and the internal analyses advanced for it. 3 introduces and discusses the processing theory of Hawkins (2004); 4 recasts two of Hawkins's metrics in computational terms and thus derives a statistical FOFC. Section 5 concludes.

1   EXPLANATIONS IN SYNTAX

Much recent work within the generative tradition has attempted to go beyond explanatory adequacy as defined by Chomsky (1964:63) and adopt something closer to the functionalist notion of explanation. Hauser, Chomsky and Fitch (2002) pose the question of "to what extent the computational system [of natural language] is optimal" in the sense of meeting natural conditions of efficiency:

(1)   "To the extent that this can be established, we will be able to go beyond the ... accomplishment of finding the principles of [FL], to an understanding of why the faculty follows these particular principles and not others." (2002:1578)

This challenge has been taken up by much work within the Minimalist Program. In an influential paper, Chomsky (2005) argues that the growth of language is affected by three factors:

(2)   "1. Genetic endowment [...]
    2. Experience, which leads to variation [...]
    3. Principles not specific to the faculty of language." (2005:6)

The third factor, which includes "principles of efficient computation", is a new introduction to the Chomskyan enter Importantly, these three factors are active on the level of "the individual" (Chomsky 2005:6), meaning that third-factor prin may be involved not only in the evolution of UG but also in the development of language in each human brain.

For present purposes, however, it is enough to consider the possibility that principles of UG may have emerge reasons of computational efficiency. If so, there may in fact be *two* valid explanations for a given linguistic phenomenon: a factor explanation in terms of UG principles, and a third-factor external explanation illuminating why such principles may arisen. In what follows I present a potential case of this.

---

2   FOFC

FOFC has its roots in a typological tradition dating back to Greenberg (1963), who presented a number of typological universals. Although these were observational, he did tentatively posit a notion of "harmony" in which VO was harmonic with VS, NGen, NA and prepositions while OV was harmonic with the opposites (1963:98-100). This work inspired a rich field of investigation (e.g. Hawkins (1983); Dryer (1992)). However, as Dryer (1992:fn17) recognised, disharmonic languages outnumber harmonic languages, so no absolute generalisations can be made on this basis. More recently, though, it has been shown that a different generalisation holds among "disharmonic" languages:

(3)   *The Final-Over-Final Constraint (FOFC)*
         If α is a head-initial phrase and β is a phrase immediately dominating α, then β must be head-initial. If α is a
         head-final phrase, and β is a phrase immediately dominating α, then β can be head-initial or head-final.
         (Holmberg 2000:124)

Biberauer, Holmberg and Roberts (BHR) (2007, 2008) present much empirical evidence for this: the order V-O-Aux example, is absent in Finnish, Germanic and Basque (2008:3-5). However, although (3) is a strong statistical generalis exceptions do exist. German, for example, features head-initial DPs and PPs dominated by head-final VPs (BHR 2007:9):

(4)   Johann hat [$_{VP}$ [$_{DP}$ den Mann] gesehen]
         John   has   the man   seen
         'John has seen the man.'

BHR therefore reformulate FOFC as in (5):

(5)   If a phase-head PH has an EPP-feature, then all the heads in its complement domain *from which it is non-*
         distinct in categorial features must have an EPP-feature. (BHR 2008:13)

In this model, the underlying order is always head-complement, and EPP-features derive complement-head linear order. The asymmetry in (3) is captured by the fact that the reverse requirement is not made: if a phase-head does not have an EPP-feature, heads in its domain may still have one. The clause in italics is necessary to account for the non-applicability of FOFC in cases like (4). (5) does not follow directly from any previously assumed theoretical apparatus; however, as BHR note, it is in the spirit of Chomsky's (2006:15) suggestion that phase-heads transmit their Agree-features to heads in their domain.

       This account is explanatorily adequate, then, and seems to make the correct predictions empirically. However, it makes no reference to third factors or interface properties: Chomsky relates the transmission of Agree-features to the necessity of maintaining the A-A' distinction at the conceptual-intentional interface (2006:10-11), but this cannot obviously motivate (5). Within the Minimalist Program, this lack of external motivation is problematic: as Sheehan (2008:6) notes, "if FOFC stems from some formal property of language, ... according to the guidelines of the Minimalist Program, it must be that it is derived from assumptions based on virtual conceptual necessity."

3   PROCESSING EFFICIENCY: AN EXTERNAL EXPLANATION?

It remains worthwhile, then, to look for external factors that could lead to a better understanding of FOFC. One such may be the processing theory proposed in Hawkins (1990, 1994, 2004, 2007). The key ideas of this theory are given in (6).

(6)   *Early Immediate Constituents* (EIC)
         The human processor prefers linear orders that maximise PCDs (by maximizing their IC-to-nonIC [or IC-to-
         word] ratios), in proportion to the minimization difference between competing orders.

         Phrasal Combination Domain (PCD)
         The PCD for a mother node M and its I(mmediate) C(onstituent)s consists of the smallest string of terminal
         elements ... on the basis of which the processor can construct M and its ICs. (Hawkins 2004:107)

In other words, PCDs that are as short as possible will be preferred. PCD length alone does not predict a preference between "disharmonic" orders, however. In Hawkins (1990, 1994), an on-line 'aggregate' ratio for PCDs is calculated; this incorporates a preference for early ICs within the PCD itself, which serves to derive a statistical FOFC (see Sheehan (2008:7) for discussion). More recent versions of the theory (Hawkins 2004, 2007) abandon aggregates, instead subsuming

EIC under Minimize Domains (MiD) (2004:32), keeping the basic metric untouched, and positing another principle, Maximize On-line Processing (MaOP) (2004:51). This is quantified as follows:

(7)   *On-line property to ultimate property ratio*
       The OP-to-UP ratio is calculated at each word X within a connected set of words ... The cumulative number of properties assignable at each X is divided by the total number of properties to be assigned in the connected set in the ultimate representation of each structure, and the result is expressed as a percentage ... The higher these on-line percentages, the more efficient ... since more properties are assigned earlier. (Hawkins 2004:55)

This principle interacts with MiD/EIC to predict favoured orderings. EIC prefers harmonic orders such as VO & CompS and OV & SComp. OP-to-UP, on the other hand, will always prefer CompS, as this assigns the property of belonging to a dependent clause earlier. The two principles thus have the same preferences in VO languages, but differ for OV. Neither principle favours the unattested (Dryer 1992:102; Zwart 2007), FOFC-violating VO & SComp. The same argument can be extended to (e.g.) the DP, and we thus derive a statistical FOFC by a different route.

     As an explanation for FOFC, this account is problematic in some respects. Sheehan observes (2008:17-18) that articles (D) are verb-patterners according to Dryer (1992:103-4), but OV & D-NP is no less common than VO & NP-D, and so there is no evidence that FOFC holds for DP complements of V (e.g. (4)). All other things being equal, OP-to-UP prefers DP-initial articles in order to assign the property of attachment to DP earlier; this prediction does not appear to hold. *Prima facie*, the non-existence of FOFC effects between DP and V etc. is therefore problematic for this theory (although see Hawkins (2002:134-7) and section 4 of this squib). As the internal account only holds between non-categorially-distinct heads, it does not encounter this problem.

     A second issue problematic for this account is that the effects of FOFC are absolute for at least some categories. As mentioned above, no convincing example of VO & SComp can be found cross-linguistically (Zwart 2007). EIC and OP-to-UP are quantitative and relative, making no claim of absolute non-occurrence. For such cases a prohibition within UG such as (5) is more satisfactory. On the other hand, for some categories, e.g. V and PP, a statistical FOFC effect seems to hold despite a few robust counter-examples (e.g. Tigré; Sheehan 2008:16), which supports a quantitative approach, as (5) does not capture this.

     Thirdly, FOFC appears to hold of heads that are not spelled out: in some OV languages, clausal extraposition to avoid FOFC violation occurs obligatorily even where an initial C is unrealised. This is not predicted by Hawkins's theory, as his syntax does not include "empty" categories and the EIC/OP-to-UP metrics therefore would not disprefer sentences without extraposition.

     This leads on to the most fundamental problem with a processing account, namely that it is unclear how parsing (dis)preference leads to a structure being disfavoured/prohibited in the grammar. This Performance-Grammar Correspondence Hypothesis is central to Hawkins's theory (2004:3), but the mechanism is never made explicit. Introspection on the part of the speaker/learner is required in order to translate an instance of processing dispreference into a parameter setting or bias against production; given the (tacit) nature of linguistic knowledge it seems unlikely that such introspection could occur. At the least we would need a theory of *how* this introspection works, which Hawkins does not provide.

     OP-to-UP is particularly problematic as a processing principle, since it involves look-ahead: in order for the ratio to be calculated at each word, the total number of properties of the structure must be known, yet if the structure has not yet been fully processed this is impossible. The only way for (7) to be instantiated mentally would be for evaluation to take place *after* the sentence had been fully parsed, defeating the object of an on-line metric.

     We can conclude that a processing account of FOFC, while appealing in that it provides external motivation for the typological facts, is not theoretically adequate on its own.

4   RELEVANCE OF A QUANTITATIVE APPROACH

Thus far I have discussed Hawkins's "external" approach and BHR's "internal" approach as if they were competing explanations for the data. However, this is an oversimplification. The internal approach outlined in section 2 is explanatorily adequate; however, it ultimately stipulates a property of UG, and within Minimalism it is desirable to derive stipulations from virtual conceptual necessity. An external approach such as that described in section 3 is arguably more deeply explanatory, and its quantitative metrics make many correct predictions; however, as I have shown, it cannot be taken to supplant a competence-based account of the data. Is there any way to combine the two explanations?

     Mobbs (2008) is an interesting critique of Hawkins (2004), reformulating its key principles as third-factor principles of computational efficiency of the kind suggested by Chomsky (2005). He illustrates how aspects of such principles are *already* instantiated in Minimalist theories of grammar: for instance, it is suggested that the LCA of Kayne (1994) is an instantiation of (a computational version of) MaOP (2008:37-8). He does not, however, discuss the quantitative metrics

used to calculate EIC or OP-to-UP ratios. Since these provide the basis for the well-supported empirical predictions of Hawkins's theory, including the statistical FOFC, they are arguably its most appealing and valuable feature. In this section I recast these two metrics so that they are compatible with a UG-based FOFC.

Much of Hawkins's work is neutral on the issue of production vs. comprehension: "if EIC can be systematically generalized from a model of comprehension to a model of production... then so much the better" (1994:427). Data from production also seems to confirm EIC benefits for the speaker (Hawkins 2004:106). We have seen in 3 that conceptualising the metrics as measuring *processing* preferences is problematic; I will therefore follow Mobbs (2008:12) in making no reference to processing, instead concentrating on computation/production.

A consequence of this is that we must abandon Hawkins's representational model of syntax, which allows ternary branching and rejects empty categories: it is not clear how derivation might proceed in such a model, whereas this has been studied extensively within Minimalism. This step immediately resolves the problem of FOFC holding of unrealised heads mentioned in 3.

EIC is simple to reformulate in these terms. Given binary branching, each PCD (defined as in (6)) may only contain two ICs. Intuitively, and in Hawkins's model, one of these ICs is always the head of the phrase, e.g. V for VP. The other IC is the head of that phrase's complement, e.g. P of PP in (7). Minimising the PCD, then, means minimising the amount of material that intervenes in linear order between a head and the head hierarchically below it. One interpretation of this is that worse ratios are associated with a higher working memory load or *space complexity* (Mobbs 2008:13). We are now in a position to recast EIC:

(8)   *EIC$_1$*
        Human computation prefers linear orders that minimise the number of terminals intervening between a head
        and the head hierarchically below it.[1]

As two heads are always involved, the metric equivalent to the original ratio is:

(9)   2/(number of terminal nodes in PCD)

expressed as a percentage. There are two ways of creating an order that is "optimal" with respect to (8) and (9): consistent head-finality, or consistent head-initiality. We thus derive harmony.

It is, however, the OP-to-UP metric that gives rise to FOFC effects in Hawkins (2004). This is much more difficult to calculate than EIC due to the breadth of the notion of "property": syntactic categories, mother-daughter attachments and "various argument structure relations of combination and dependency" are all classed as properties (2004:56). Evidently, different theoretical assumptions will lead to major differences in ratios. For this reason I will not pursue calculations here, instead concentrating on the concept.

(7) is more suitable for instantiation in production than in processing, because, as discussed in section 3, it involves look-ahead to the total number of properties in the structure: the number of properties "assigned" at each word X is divided by this total. The efficiency of the structure can then be calculated by averaging all these percentages, as for EIC-aggregate ratios.[2] The need to know the ultimate number of property assignments in advance means that this metric is perhaps best viewed as a measure of optimality at linearisation. Following Chomsky (2001:12-13) and BHR (2007, 2008) I will assume that strong phases are the object transferred from narrow syntax to the phonological component. At PF, then, the syntactic object has already been constructed and hence a certain number of properties/relations are already in working memory. At this point it is desirable to 'offload' these properties as soon as possible, in order to reduce *time complexity* as defined by Mobbs (2008:13). This translates to a preference for items associated with more properties to precede items with fewer properties at linearisation, and this preference can be quantified using a computational variant of the OP-to-UP metric:

(10)   *OP-to-UP$_1$*
        The ratio is calculated at each terminal X within a phase. The cumulative number of properties offloadable
        at each X is divided by the total number of properties to be offloaded in the phase, and the result is
        expressed as a percentage. These percentages are averaged to quantify the computational efficiency of the
        phase in linearisation.

The metrics have been saved, then, with a little reconceptualisation. EIC$_1$ prefers harmonic ordering of heads; the OP-to-UP$_1$ metric generally prefers head-initiality. FOFC itself (3) is derived when we consider the potential effect of these

---

[1]   Unlike Mobbs I have not removed reference to weighting/preference, because I do not believe that third factors are "expected to serve in a discrete, formal grammar" (2008:12). As Chomsky (2005:6) makes clear, such principles are by definition *distinct* from UG, although they enter into the growth of grammar. In any case, quantitative preference is implicit where relative terms such as "minimisation" are used.

[2]   Hawkins uses the metric to compare structures directly, by calculating the difference between percentages at each word and then summing these differences (2004:56-7), but it is difficult to imagine comparison taking place in FL. Averaging yields similar results.

metrics in syntactic derivation. If trees are constructed bottom-up, then orders dispreferred by *both* EIC_1 and $OP\text{-to-}UP_1$ arise only through Merge of an EPP-feature-bearing head with a phrase whose head does not bear such a feature. This operation would create the structure ruled out by the constraint in (3), in which a head-final phrase immediately dominates a head-initial phrase, e.g. V-O-Aux. (The opposite, "anti-FOFC" Merge exemplified by Aux-O-V will usually create structures preferred by $OP\text{-to-}UP_1$.) The effect created by the interaction of EIC_1 and $OP\text{-to-}UP_1$ in computation is thus equivalent to (3), albeit quantitative rather than absolute.

It is an open question how such third-factor principles would affect any given grammar. One possibility is via an acquisition preference such as the Superset Bias (Boeckx 2008): "Strive for parametric value consistency among similar parameters", with the relevant microparameters being the presence/absence of EPP-features on heads (cf. also Roberts 2007:273). Alternatively a least effort strategy in production could reduce the frequency of computationally inefficient structures in the PLD, thereby reducing the likelihood of acquisition of "inefficient" parameter values; for example, it is intuitively obvious that "center-embedding" (Kuno 1974) sentences such as 'The cheese the mouse the cat chased ate was rotten' are difficult to produce as well as to process. Either position is consistent with the holistic functionalism advocated by Newmeyer (2002), but it is not necessary to commit to either. As we have observed, some structures are disallowed in all languages, which indicates that something like (5) is required as a UG principle. The contribution of the present squib is to argue that third-factor computational preferences, such as those presented in this section, may underlie its emergence.

This approach explains (3), but provides no immediate insight into the categorial distinction built into (5), necessary to account for exceptions. It may be possible to account for these in a precise formulation of $OP\text{-to-}UP_1$: Hawkins (2002:134) argues that while the object depends upon V for e.g. case assignment and construction of VP, the verb itself is dependent on the object for selection of its subcategorisation frame. If the number of properties offloaded by V and by a D head of its DP complement is equal, then $OP\text{-to-}UP_1$ will be neutral as to which should precede; depending on how we define 'property', this could explain cases (such as (4)) that violate (3). I leave this matter for future research.

5   CONCLUSION

I have argued that FOFC as given in (5) is not obviously motivated by virtual conceptual necessity. In the spirit of Mobbs (2008), I reformulated/reconceptualised some of the metrics of Hawkins (2004) as third-factor principles of computational efficiency compatible with a UG-based FOFC, and suggested that the conceptual justification of a formal principle such as (5) might be derived from such third factors.

Obviously this approach raises a number of questions. Due to differing syntactic assumptions, the reformulated metrics make empirical predictions different to those of the originals; whether they fare better or worse is a matter for empirical research. (5) also includes a categorial distinction not obviously explained by computational efficiency. Finally, it is worth investigating exactly how third-factor principles might be implicated in the evolution of FL, although the question seems virtually unanswerable at present.

Despite these questions, I hope that the basic insight of this paper – that FOFC may have emerged for reasons of efficient computation – will remain worthy of consideration.

REFERENCES

Biberauer, T., Holmberg, A., and Roberts, I.G. (2007) 'Disharmonic word-order systems and the Final-over-Final Constraint (FOFC).' In A. Bisetto & F. Barbieri (eds.) *Proceedings of XXXIII Incontro di Grammatica Generativa*. pp.86-105.

Biberauer, T., Holmberg, A., and Roberts, I.G. (2008) 'Linearising disharmonic word orders: the Final-over-Final Constraint.' In K-A. Kim (ed.) *Perspectives on Linguistics in the 21st Century*. pp.301-318.

Boeckx, C. (2008) *Approaching parameters from below*. ms. University of Harvard. http://ling.auf.net/lingBuzz/000641. Last accessed: 17th December 2008.

Chomsky, N. (1964) 'Current issues in linguistic theory.' In J.A. Fodor & J.J. Katz (eds.) *The structure of language*. Englewood Cliffs, NJ: Prentice-Hall, pp.50-118.

Chomsky, N. (2001) 'Derivation by phase.' In M. Kenstowicz (ed.) *Ken Hale: a life in language.* Cambridge, Mass.: MIT Press, pp.1-52.

Chomsky, N. (2005) 'Three factors in language design.' *Linguistic Inquiry* 36:1-22.

Chomsky, N. (2006) On phases. In R. Freidin, C. Otero & M. Zubizaretta (eds.) *Foundational issues in linguistic theory*. Cambridge, Mass.: MIT Press, pp.133-166.

Dryer, M.S. (1992) 'The Greenbergian word order correlations.' *Language* 68:81-138.

Greenberg, J. (1963) 'Some universals of grammar with particular reference to the order of meaningful elements.' In J. Greenberg (ed.) *Universals of grammar*. Cambridge, Mass.: MIT Press, pp.73-113.

Hauser, M.D, Chomsky, N., & Fitch, W.T. (2002) 'The faculty of language: what is it, who has it, and how did it evolve?' *Science* 298:1569-1579.

Hawkins, J.A, (1983) *Word order universals*. New York: Academic Press.

Hawkins, J.A. (1990) 'A parsing theory of word order universals.' *Linguistic Inquiry* 21:223-261.

Hawkins, J.A. (1994) *A performance theory of order and constituency*. Cambridge: CUP.

Hawkins, J.A. (2002) 'Symmetries and asymmetries: their grammar, typology and parsing.' *Theoretical Linguistics* 28:95-149.

Hawkins, J.A. (2004) *Efficiency and complexity in grammars*. Oxford: OUP.

Hawkins, J.A. (2007) 'Processing typology and why psychologists need to know about it.' *New Ideas in Psychology* 25:87-107.

Holmberg, A. (2000) 'Deriving OV order in Finnish.' In P. Svenonius (ed.) *The derivation of VO and OV*. Amsterdam: John Benjamins, pp.123-152.

Kayne, R. (1994) *The antisymmetry of syntax*. Cambridge, Mass.: MIT Press.

Kuno, S. (1974) 'The position of relative clauses and conjunctions.' *Linguistic Inquiry* 5:117-36.

Mobbs, I. (2008) *'Functionalism', the design of the language faculty, and (disharmonic) typology.* ms. University of Cambridge. http://ling.auf.net/lingBuzz/000680. Last accessed: 7th December 2008.

Newmeyer, F.J. (2002) 'Where is functional explanation?' In M. Andronis, C. Ball, H. Elston, & S. Neuvel. (eds.) *Papers from the Thirty-Seventh Meeting of the Chicago Linguistic Society*. Chicago: Chicago Linguistic Society, pp.99-122.

Roberts, I.G. (2007) *Diachronic Syntax*. Oxford: OUP.

Sheehan, M. (2008) *The final-over-final constraint: trend or principle?* ms. Newcastle University.

Zwart, J.-W. (2007) 'Relevance of typology to Minimalist inquiry.' *Lingua* to appear. http://www.let.rug.nl/zwart/docs/lingua2reg.pdf. Last accessed: 31st December 2008.

George Walkden

Linguistics Department
Sidgwick Avenue
University of Cambridge
Cambridge
CB3 9DA
United Kingdom

gw249@cam.ac.uk