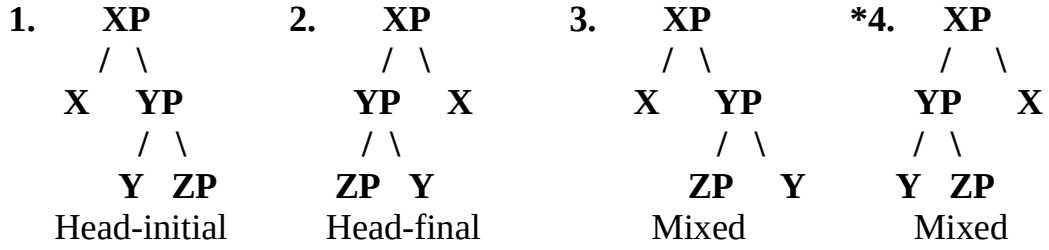
Disharmonic Word Order from a Processing Typology Perspective

John A. Hawkins, U of Cambridge RCEAL & UC Davis Linguistics

[A] Introduction



3 and 4 are 'inconsistent' or 'disharmonic' word orders in the language typology research tradition (Greenberg 1966, Hawkins 1983, Dryer 1992), 1 and 2 are consistently and harmonically head-initial and head-final respectively. Within formal grammar a proposal has recently been made for a different partitioning that distinguishes the mixed type in 4 from the other three:

5. **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.

The FOFC rules out 4, and permits 1-3. The FOFC is derived from principles of Minimalist Syntax (Chomsky 2000, Kayne 1994, Biberauer, Holmberg & Roberts 2007, 2008).

From a typological perspective the FOFC looks, prima facie, like it's not quite right: languages with *4 are generally dispreferred, occasionally unattested (i.e. it's too strong); while languages with 3 appear to be similarly dispreferred, occasionally unattested (too weak); 1 and 2 are fully productive.

Some Greenbergian word order correlations (Hawkins 1983, Dryer 1992)

6. a. vp[went pp[to the movies]] (1) b. vp[pp[the movies to] went] (2)
c. vp[went pp[the movies to]] (3) d. vp[pp[to the movies] went] (*4)

6'. a. vp[V pp[P NP]] = 161 (41%) b. vp[pp[NP P] V] = 204 (52%)
c. vp[V pp[NP P]] = 18 (5%) d. vp[pp[P NP] V] = 6 (2%)
Preferred (6a)+(b) = 365/389 (94%) [Data from Dryer's 1992 sample]

. a. pp[P np[N Possp]] = 134 (40%) (1) b. pp[np[Possp N] P] = 177 (53%) (2)
c. pp[P np[Possp N]] = 14 (4%) (3) d. pp[np[N Possp] P] = 11 (3%) (*4)
Preferred (7a) + (b) = 311/336 (93%) [Data from Hawkins 1983]

Typologists and formal grammarians can help each other identify the precise cross-linguistic regularities in this area (Hawkins 1985). At an explanatory level they can both benefit from considering the possible role of processing in shaping these regularities (Hawkins 1994,2004).

[B] The Processing Typology Research Programme

8. Performance-Grammar Correspondence Hypothesis (PGCH)
Grammars have conventionalized syntactic structures in proportion to their degree of preference in performance, as evidenced by patterns of selection in corpora and by ease of processing in psycholinguistic experiments.

The PGCH is an attempt to make sense of cross-linguistic variation in terms of principles of performance. It makes predictions for occurring and non-occurring lg types, for frequent and less frequent ones. It can also motivate many of the stipulated principles of formal grammar.

Heads = a subset of mother node constructing categories (Hawkins 1994:ch.6)

9. **Mother Node Construction** (Hawkins 1994:62; cf. Kimball's 1973 New Nodes)
In the left-to-right parsing of a sentence, if any word of syntactic category C uniquely determines a phrasal mother node M, in accordance with the PS rules of the grammar, then M is immediately constructed over C.

10. **Immediate Constituent Attachment** (Hawkins 1994:62)
In the left-to-right parsing of a sentence, if an IC does not construct, but can be attached to, a given mother node M, in accordance with the PS rules of the grammar, then attach it, as rapidly as possible. Such ICs may be encountered *after* the category that constructs M, or *before* it, in which case they are placed in a look-ahead buffer.

Why is it that certain linear orderings of words are preferred over others in performance and in grammars? Because there are principles of processing efficiency that motivate the preferences. E.g. the adjacency of V and P in (6ab) guarantees the smallest possible string of words for construction of VP and of PP, and for attachment of V and PP to VP as sister ICs. Non-adjacency of heads in (6cd) is less efficient for phrase structure processing.

Hypothesis: the construction of phrases and the recognition of their combinatorial and dependency relations prefers the smallest possible string of words for processing (the principle of Early Immediate Constituents, Hawkins 1994); more generally the processing of all syntactic and semantic relations prefers minimal domains, cf. also Gibson's (1998) "locality".

11. **Minimize Domains** (MiD) [Hawkins 2004]
The human processor prefers to minimize the connected sequences of linguistic forms and their conventionally associated syntactic and semantic properties in which relations of combination and/or dependency are processed. The degree of this preference is proportional to the number of relations whose domains can be minimized in competing sequences or structures, and to the extent of the minimization difference in each domain.

Structures 1 and 2 = optimal by MiD: two adjacent words suffice for construction of the mother XP (projected from X) and for construction of YP (projected from Y) and its attachment to XP as a sister of X. Structures 3 and 4 = less efficient: more words must be processed for construction and attachment.

MiD predicts Head Adjacency and the Head Ordering Parameter (cf. Newmeyer 2005:43). One and the same principle can explain both the preferred conventions of grammars as well as preferred structural selections in performance in languages and structures in which speakers have a choice, cf. Hawkins (1994, 2004) for summary of performance data from many lgs.

MiD can also explain why there are two highly productive mirror-image types, head-initial and head-final languages, i.e. 1 and 2. They are equally efficient. Structures 3 and 4 are not as efficient and both are significantly less productive.

A second interacting principle:

12. **Maximize On-line Processing** (MaOP) [Hawkins 2004]
The human processor prefers to maximize the set of properties that are assignable to each item X as X is processed, thereby increasing O(n-line) P(roperty) to U(ltimate) P(roperty) ratios. The maximization difference between competing orders and structures will be a function of the number of properties that are unassigned or misassigned to X in a structure/sequence S, compared with the number in an alternative.

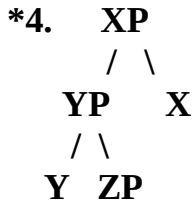
[C] Structures 1-4 and the Timing of Phrasal Constructions and Attachments

1. X constructs XP, then Y constructs YP at the next word & YP is immediately attached left as daughter to mother XP. (Processing of ZP follows.)
2. (Processing of ZP first.) Y constructs YP, then X constructs XP at the next word & YP is immediately attached right as daughter to mother XP. (NB! The attachment of YP follows its construction by 1 word)

3. X constructs XP, then after processing ZP Y constructs YP & YP is attached left to mother XP, possibly several words after construction of XP (Delayed Assignment of Daughter YP to XP)
4. Y constructs YP first, then after processing ZP X constructs XP & YP is attached right to mother XP, possibly several words after construction of YP (Delayed Assignment of Mother XP to YP)

	MiD	MaOP
Structure 1	optimal	adjacent words for XP & YP construction & attachments
Structure 2	optimal	adjacent words for XP & YP construction & attachments
Structure 3	non-optimal	non-adjacent … Delayed Daughter YP assignment to XP
Structure *4	non-optimal	non-adjacent … Delayed Mother XP assignment to YP

[D] Processing Typology Predictions for Structure *4



Delayed assignment of mother XP to daughter YP, i.e. **No Mother On-line for YP** for several words of processing!

(a) Limit productivity of *4 compared with 2 as basic orders (keeping X final)
- (i) vp[np[N Possp] V] vs. vp[np[Possp N] V] = 9.7% genera (12/124) Dryer 1992
- (ii) vp[pp[P NP] V] vs. vp[pp[NP P] V] = 6.1% genera (7/114) Dryer 1992
- (iii) tp[vp[V NP] T] vs. tp[vp[NP V] T] = 10% genera (4/40) Dryer 1992
- (iv) np[cp[C S] N] vs. np[cp[S C] N] = 0 Lehmann 1984

(b) Limit productivity of *4 compared with 1 as basic orders (keeping Y initial)
- (i) vp[np[N Possp] V] vs. vp[V np[N Possp]] = 16% genera (12/75) Dryer 1992
- (ii) vp[pp[P NP] V] vs. vp[V pp[P NP]] = 9.1% genera (7/77) Dryer 1992
- (iii) tp[vp[V NP] T] vs. tp[T vp[V NP]] = 12.5% genera (4/32) Dryer 1992
- (iv) np[cp[C S] N] vs. np[N cp[C S]] = 0 Lehmann 1984

Prediction: the more structurally complex YP is, the more it will be dispreferred in *4, e.g. CP worse than NP or PP, cf. (iv).

(c) Non-rigid OV vs. rigid OV languages

Non-rigid OV: lgs with basic OV that combine pre- and post-verbal phrases in VP (Greenberg 1966). Such lgs are predicted here to be those that combine Y-

initial YP with X-final XP, i.e. type *4, and they are further predicted to postpose YP to right of V, in proportion to the complexity of YP, creating alternations with structure 1.

E.g. obligatory extraposition of vp[cp[C S] V] => vp[V cp[C S]] in Persian and German and other such lgs (Dryer 1980, Hawkins 1990):

13. a. *An zan cp[ke an mard sangi partab kard] mi danat (Persian)
the woman that the man rock threw CONT knows
'The woman knows that the man threw a rock'
b. An zan mi danat cp[ke an mard sangi partab kard]

78% (7/9) OV genera in WALS with prepositions (rather than postpositions) = non-rigid OV rather than rigid, and PPs regularly follow V in these lgs converting *4 into 1 (Hawkins 2008

73% (8/11) OV genera in WALS with np[N Possp] (i.e. postnominal rather than pronominal genitives) = non-rigid OV rather than rigid, and NPs regularly follow V in these lgs (ibid)

Rigid OV: lgs with basic OV in which V is final in VP and sisters precede. Such lgs are predicted here to combine X-final XP (i.e. OV) with Y-final YP.

96% (47/49) rigid OV genera in WALS have postpositions (rather than prepositions), i.e. vp[pp[NP P] V] (Hawkins 2008, Haspelmath, Dryer, Gil & Comrie 2005)

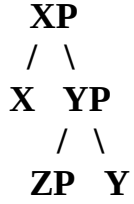
94% (46/49) rigid OV genera in WALS have vp[np[Possp N] V] (i.e. prenominal rather than postnominal genitives) (Hawkins 2008, Haspelmath, Dryer, Gil & Comrie 2005)

(d) Keep YP in situ in *4 but extrapose (out of) ZP, shortening YP

14. a. Ich habe vp[np[den Lehrer cp[der das Buch geschrieben hat]] gesehen] (German)
I have the teacher who the book written has seen
'I have seen the teacher who wrote the book'
b. I habe vp[np[den Lehrer] gesehen] cp[der das Buch geschrieben hat] (Hawkins 2004)

[E] Processing Typology Predictions for Structure 3

3.



Delayed assignment to a constructed mother XP of a daughter YP, i.e. **No Daughter On-line for XP** for several words of processing.

(a) Limit productivity of 3 compared with 1 as basic orders (keeping X initial)

(i)	vp[V np[Possp N]] vs. vp[V np[N Possp]]	= 32% (30/93) genera	Dryer 1992
(ii)	vp[V pp[NP P]] vs. vp[V pp[P NP]]	= 14.6% (12/82) genera	Dryer 1992
(iii)	tp[T vp[NP V]] vs. tp[T vp[V NP]]	= 9.7% (3/31) genera	Dryer 1992
(iv)	np[N cp[S C]] vs. np[N cp[C S]]	= v. few, if any	Lehmann 1984
(v)	vp[V cp[S C]] vs. vp[V cp[C S]]	= 0	Hawkins 1990

(b) Limit productivity of 3 compared with 2 as basic orders (keeping Y final)

(i)	vp[V np[Possp N]] vs. vp[np[PosspN] V]	= 21.1% (30/142) genera	Dryer 1992
(ii)	vp[V pp[NP P]] vs. vp[pp[Possp N] V]	= 10.1% (12/119) genera	Dryer 1992
(iii)	vp[T vp[NP V]] vs. tp[vp[NP V] T]	= 7.7% (3/39) genera	Dryer 1992
(iv)	np[N cp[S C] vs. np[cp[S C] N]	= v. few, if any	Lehmann 1984
(v)	vp[V cp[S C]] vs. vp[cp[S C] V]	= 0	Hawkins 1990

Prediction: the more structurally complex ZP is, the more it will be dispreferred in 3, e.g. S is worse than NP or PossP in (iv) and (v).

(c) Construct YP early in advance of Y thru alternative constructors in ZP

E.g. preposing of non-nominative case-marked pronouns and full NPs in German VP serves to construct VP at or near left periphery by Grandmother Node Construction (Hawkins 1994:361), e.g. in tp[T vp[NP … V]]

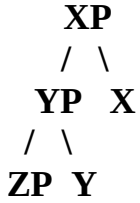
15. Ich tp[habe vp[ihn [noch einmal] gesehen]
I have him (+Acc) once again seen
'I have seen him once again'

(d) Avoid on-line ambiguity between YP and ZP or nodes dominated by ZP

Both complexity of S and potential on-line misassignments (/garden paths) can explain the non-occurrence of vp[V cp[S C]] in (v), cf. the on-line ambiguity of *I believe the clever student wrote* …, disambiguated only at *wrote*.

[F] Processing Typology Predictions for Structure 2 (Head Finality)

2.



2 is optimal for MiD (11), but YP is constructed at Y and must then wait one word for attachment to XP until X has constructed XP. I.e. **No Mother On-line for YP** for one word of processing. Head-initial lgs (1) construct YP and attach it to XP simultaneously, with no processing delay.

(a) Fewer free-standing X words following Y, instead more X affixes on Y constructing YP and XP simultaneously at Y (the former through MNC (9), the latter through Grandmother Node Construction from an X affix on Y, Hawkins 1994:361)

E.g. the asymmetry between prepositions in head-initials lgs and postpositions in head-final, i.e. pp[P NP] vs. pp[NP P]. Postpositions are not as productive in head-final lgs as prepositions are in head-initial:

> many head-final lgs have very limited postpositions, sometimes just one or two; many lgs with strong head-final characteristics have no free-standing postpositions, but only suffixes with adposition-type meanings and a larger class of NPs bearing rich case features, 29% (19/66) in the sample of Tsunoda, Ueda & Itoh (1995:757); prepositional lgs retain free-standing prepositions productively (cf. Hall 1992)

Complementizers, i.e. free-standing words that construct subordinate clauses (vs participial and other subordinate clause indicators affixed to verbs) are much less productive in head-final than in head-initial lgs:

> Of lgs with free-standing complementizers, 74% (140) occur (initially in CP) in VO lgs, i.e. structure 1, just 14% (27) occur (finally) in OV lgs, i.e. structure 2 (and 12% (22) initially in OV), cf. Dryer (2007).
>
> Adding affixes to verbs that indicate subordinate clause status in OV lgs means that both S and subordinate status are constructed simultaneously on the last word of the subordinate clause.

(b) Avoid additional constructors of phrasal nodes in OV, but not VO, lgs

Assume (controversially given the DP theory) that definite articles construct NP, just like N or Pro and other categories uniquely dominated by NP. If so either N or Art can construct NP immediately on its left periphery and provide efficient and minimal "phrasal combination domains" (PCDs) in VO lgs. Art-initial is especially favored when N is not initial in NP.

16. vp[V np[N ... Art ...]
    vp[V np[Art ... N ...]
        |------|

In OV languages any additional constructor of NP will lengthen these processing domains, whether it follows or precedes N, by constructing the NP early and extending the processing time from the construction of NP to the processing of V. Additional constructors of NP are therefore inefficient in OV orders.

17. [[... N ... Art]np V]vp
    [[... Art ... N]np V]vp
        |--------------|

18.	Def word distinct from Dem	No definite article	[WALS data]
Rigid OV	19% (6)	81% (26)	
VO	58% (62)	42% (44)	

This same consideration provides a further motivation for the absence of free-standing complementizers in head-final languages. Complementizers can shorten PCDs when they precede V in VO lgs, by constructing subordinate clauses on their left peripheries (*John knows* [***that*** *he is sick*]), but they will lengthen PCDs in OV lgs, compared with projections from V alone, whether they are clause-initial or clause-final.

(c) Reduce left-branching YP and ZP phrases

E.g. Lehmann (1984:168-73) observes that prenominal relative clauses = significantly more restricted in their syntax and semantics than postnominal rels: greater nominalization (or non-sentential properties); less tolerance of appositive interpretations. The former results in fewer tense, aspect and modal forms, non-finite verbs, less embedding, conversion of subject to genitive, etc.

[G] Conclusions

(a) These typological patterns suggest that the FOFC (as formulated in 5) is not quite capturing the right generalization: it appears to be too strong (structure *4 is generally dispreferred, occasionally unattested), and too weak (structure 3 is also dispreferred, occasionally unattested).

(b) Typologists need the greater precision of in-depth analysis for their languages sampled, as provided by formal syntax, in order to determine what exactly the cross-linguistic patterns are, how best to formulate them, what the relevant syntactic categories are, etc.

(c) Conversely formal syntacticians need to heed the fact that structure 3 looks almost as bad in these typological correlations as *4. It is misleading of them to suggest that all of 1-3 are common, with *4 the only violation.

(d) Typologists need a more sophisticated theoretical basis, and more explanatory theories, for their cross-linguistic correlations. The goal of the Processing Typology research programme (section [B]) is to provide one: it brings an independent body of evidence from language performance and psycholinguistics (esp. processing) to bear on cross-linguistic grammatical conventions and parameters. The central hypothesis is the PGCH (8): grammars have conventionalized syntactic structures in proportion to their degree of preference in performance.

(c) The rich theoretical apparatus of generative syntax is subtle and its descriptive coverage is impressive. But much of this apparatus is stipulated, and the appeal to an innate UG is largely speculation, and increasingly controversial (cf. the papers in Christiansen, Collins & Edelman 2009). Independent evidence from performance in diverse languages is growing meanwhile, and the preferences and dispreferences in structural selections in performance (in lgs with choices) are being shown to correlate with preferences and dispreferences

in the grammatical conventions themselves, supporting the PGCH (Hawkins 1994, 2004). The stipulations of formal models can become less stipulative by shifting their ultimate motivation away from an innate UG towards (ultimately innate and neurally pre-determined) processing mechanisms, in the manner of certain constraints of Optimality Theory (Haspelmath 1999).

(d) The PGCH defines an alternative research programme and explanation for the cross-linguistic patterns that have ultimately led to the FOFC. I suggest that typologists, formal syntacticians and psycholinguists work more closely together, in order to get the facts right, and in order to develop the explanatory ideas in more detail that have been outlined in this paper. The current workshop is an excellent move in this direction. I thank the organizers for inviting me!

References

Biberauer, T., A. Holmberg & I. Roberts (2007) 'Disharmonic word-order systems and the Final-over-Final Constraint (FOFC)', in Incontro di grammatical generativa

Biberauer, T., A. Holmberg & I. Roberts (2008) 'Structure and linearization in disharmonic word orders', in Proceedings of the 26th West Coast Conference on Formal Linguistics.

Chomsky, N. (2000) 'Minimalist inquiries: the framework', in R. Martin, D. Michaels & J. Uriagereka, eds., Step by step: Essays on Minimalist Syntax in Honor of Howard Lasnik, MIT Press, Cambridge, Mass., 89-156.

Christiansen, M.H., C. Collins & S. Edelman, eds., (2009) Language Universals, OUP, Oxford.

Dryer, M.S. (1980) 'The positional tendencies of sentential noun phrases in Universal Grammar', Canadian Journal of Linguistics 25: 123-95.

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

Dryer, M.S. (2007) 'The branching direction theory of word order correlations revisited', MS, Dept of Linguistics, SUNY Buffalo.

Gibson, E. (1998) 'Linguistic complexity: Locality of syntactic dependencies', Cognition 68: 1-76.

Greenberg, J.H. (1966) Language Universals, with Special Reference to Feature Hierarchies, Mouton, The Hague.

Hall, C.J. (1992) Morphology and Mind, Routledge, London.

Haspelmath, M. (1999) 'Optimality and diachronic adaptation', Zeitschrift für Sprachwissenschaft 18: 180-205.

Haspelmath, M., Dryer, M.S., Gil, D., and Comrie, B. (eds.). (2005) The World Atlas of Language Structures (WALS), Oxford University Press, Oxford.

Hawkins, J.A. (1983) Word Order Universals, Academic Press, New York.

Hawkins, J.A. (1985) 'Complementary methods in Universal Grammar: A reply to Coopmans', Language 61:569-587.

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, CUP, Cambridge.

Hawkins, J.A. (2004) Efficiency and Complexity in Grammars, OUP, Oxford.

Hawkins, J.A. (2008) 'An asymmetry between VO and OV languages: The ordering of obliques', in G. Corbett & M. Noonan, eds., Case and Grammatical Relations: Essays in Honour of Bernard Comrie, John Benjamins, Amsterdam, 167-190.

Kayne, R. (1994) The Antisymmetry of Syntax, MIT Press, Cambridge, Mass.

Kimball, J. (1973) 'Seven principles of surface structure parsing in natural language', Cognition 2: 15-47.

Lehmann, C. (1984) Der Relativsatz, Narr, Tübingen.

Newmeyer, F.J. (2005) Possible Languages and Probable Languages, OUP, Oxford.

Tsunoda, T., S. Ueda & Y. Itoh (1995) 'Adpositions in word-order typology', Linguistics 33: 741-61