
Keywords: Performance-Grammar Correspondence Hypothesis (PGCH), processing efficiency, functionalist, formalist, linearization

1. Introduction

1.1. Hawkins' Formal Functionalist Approach

The central claim of this book (= Hawkins (2004)) is that there is a profound correspondence between performance and grammars. Hawkins formulates this argument as the Performance-Grammar Correspondence Hypothesis (PGCH) (p. 3),\(^1\) which runs as follows.

\[\text{Performance-Grammar Correspondence Hypothesis (PGCH)}\]

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.

This hypothesis formed the basis for the parsing explanation for word order universals in Hawkins (1990, 1994), and it is supported by a number of other parsing explanations for grammars that were summa-
rized in Hawkins (1994) and are presented in this book.

Although the PGCH incorporates in its definition the significance of performance factors in shaping grammars, it does not necessarily challenge the fundamental distinction between grammar (or competence) and performance, which has been one of the background assumptions of generative grammars (Chomsky (1965) and subsequent works). Hawkins states that "it is surprising to me that there has been such a lack of curiosity in formal grammar concerning the ultimate explanation for basic principles" (p. 265), and that "the problem has been the indifference of most formal linguists to performance data, i.e. to the very data that can provide a principled choice between models and lead to greater descriptive adequacy" (ibid.). Hawkins' thirst for the ultimate explanation for general principles is reflected in the many "why" questions throughout the book. For example, the following is the quotation from Chapter 9 (p. 265): "[w]hy is there a subjacency constraint? Why are some categories adjacent and others not? Why are gaps and zero forms found in some environments and not others? Why are some categories asymmetrically ordered and others not?" (Italics on "why" are mine. See sections 3 and 4 below for part of his answers to these questions and more discussion on "ultimate explanation.")

Thus, what Hawkins attempts to do in this book is to develop a formal and hence explicit theory for why grammars are the way they are, which, according to his claim, cannot be done without looking at performance data. I call Hawkins' approach to grammars a "formal functionalist" approach in the following sense. On the one hand, Hawkins' position is ultimately functionalist in that he looks at both typological and performance variation in the pursuit of functional explanations for them. On the other hand, Hawkins uses formal insights in the pursuit of functional explanations for various linguistic phenomena, and attempts to formalize the functional explanations formally enough to be testable.

1.2. Formal Functionalist and Functional Formalist Approaches

Although I agree with Hawkins that "it would be uncharitable to deny that performance has shaped at least some grammatical rules in

2 In this regard, I disagree with Vicente (2005) in his contention that "one should not blur the competence/performance dichotomy as easily as Hawkins."
some languages” (p. 267), I will argue that there are indeed grammatical rules that cannot be given (formal) functional explanations. I will identify myself as a “functional formalist,” meaning that although some grammars may well have been shaped by performance factors diachronically, the way grammars are now is independent of functional considerations. A functional formalist position is much in the spirit of the Minimalist Program (Chomsky (1995) and subsequent works), a recent representative of formalist approaches to grammars, in that it shares the Minimalist Program’s assumption that grammars are independent of, yet interact with, performance systems. One of the goals of the Minimalist Program is to find to what extent the grammatical component is determined by “interface conditions” (Chomsky (1995: 27)), i.e. the conditions imposed on the two interfaces with external performance systems.

In this review, I mainly look at linearization phenomena, which constitute the major empirical domain of Hawkins (2004). The table below summarizes the four different approaches to linearization.

(2) Formalist and functionalist approaches to linearization

<table>
<thead>
<tr>
<th>Competence (grammar)</th>
<th>Relation</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>i. Formalist (Minimalist Program)</td>
<td>Kayne’s LCA</td>
<td>$\leftarrow \rightarrow$</td>
</tr>
<tr>
<td>ii. Functional formalist (this review article)</td>
<td>head-parameter L-to-R computation</td>
<td>$\leftarrow \rightarrow$</td>
</tr>
<tr>
<td>iii. Functionalist</td>
<td></td>
<td>$\leftarrow$</td>
</tr>
<tr>
<td>iv. Formal functionalist (Hawkins (1994, 2004))</td>
<td>grammaticalized efficiency principles</td>
<td>$\leftarrow (\rightarrow)$</td>
</tr>
</tbody>
</table>

The table should be read as follows. First, in a formalist approach (i), it is basically grammatical rules or principles that determine linear order (as indicated by the shaded performance column cell). For example, the Minimalist Program adopts Kayne’s (1994) Linear Correspondence Axiom (LCA) as a phonological principle in charge of linearization.

3 The term “linearization” is used in a broad and theory-neutral sense in this paper, referring to any linguistic phenomena that have to do with linear order.
(Chomsky (1995: 334–340)). Although it is considered that grammatical and performance rules interact to derive particular linguistic phenomena, the two types of rules are independent of each other. They interact only at the two interfaces, PF and LF. My functional formalist position (ii) is a variation of the formalist position in acknowledging that there are purely grammatical principles of linearization such as head-parameter, as well as purely performance determinants of linear order such as increasing weight or end-weight. However, it is more functional in that their interaction is more direct and profound than a purely formalist approach assumes (as is indicated by the thicker arrows in the relation column). In particular, I will propose later (section 4) that the grammar has the left-to-right (L-to-R) computation similar to that of the performance system (following Phillips (1996, 2003)).

In contrast, functionalist approaches’ primary concerns are performance systems: pure functionalists (iii) would argue that any linear order variations are ultimately traceable to performance factors such as topic-comment and old-new information packaging (see Hawkins (2004: 235–240) for discussion). Hawkins’ formal functionalist position (iv), however, is more modest than purely functionalist approaches. He looks at not only performance but also grammatical (typological) order variations in searching for functional explanations for them. His research leads him to claim that grammar is largely shaped by performance factors such as computational efficiency: performance preferences are grammaticalized as rules or principles in the grammar, and the degree of grammaticalization is subject to parameterization. It is not entirely clear how much of grammar is purely grammatical (as indicated by the parentheses around the rightward arrow). (See Newmeyer (1998) and Carnie and Harley (2003) for surveys of formalist and functionalist approaches.)

In the following sections, I will first overview the main ideas and principles proposed in the book, and illustrate how particular analyses based on these principles work (section 2). Then, I will evaluate Hawkins’ performance-based approach to grammars mainly with respect to issues on linearization (section 3). Finally, I will consider theoretical issues raised by this book, and then propose a derivational, left-to-right approach to linearization, further developing my own functional formalist program (section 4).
2. Hawkins’ Efficiency Principles and Their Predictions

This section presents a chapter-by-chapter overview of the book: chapters 1–3 (2.1), chapters 4–8 (2.2) and chapter 9 (2.3).

2.1. Three Principles

In the first three chapters, Hawkins introduces the main thesis of the book, the Performance-Grammar Correspondence Hypothesis (PGCH) we saw in (1), and defines three main principles that implement the PGCH: Minimize Domains (MiD) (p. 31), Minimize Forms (MiF) (p. 38), and Maximize On-line Processing (MaOP) (p. 51).

(3) Minimize Domains (MiD)

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.

(4) Minimize Forms (MiF)

The human processor prefers to minimize the formal complexity of each linguistic form F and the number of forms with unique conventionalized property assignments, thereby assigning more properties to fewer forms. These minimizations apply in proportion to the ease with which a given property P can be assigned in processing to a given F.

(5) Maximize On-line Processing (MaOP)

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) Property 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.

These are “efficiency principles” based on the preferences of the human processor. The principles of MiD and MiF provide definitions for the efficiency slogan “express the most with the least,” and MaOP captures
the “express it earliest” intuition (p. 25, p. 49).

Let us take a closer look at each principle. The principle of MiD in (3) subsumes Hawkins’ earlier principle of Early Immediate Constituents (EIC) (Hawkins (1990, 1994)), yet is more general than EIC. (I will come back to this comparison later in 3.1.2.) The MiD principle predicts that all syntactic and semantic relations between categories will prefer minimal domains for processing, and it defines a degree of preference that is the collective product of several possible relations between categories such as “combination” and “dependency.” (See p. 18 and p. 22 for their definitions.)

With respect to how convincing an external explanation (e.g. processing-based explanation) can be, Newmeyer (1998: 127) suggests three criteria might be identified: (i) To have precise formulation, (ii) to identify a linkage between cause and effect, and (iii) to have measurable typological consequences, and argues that EIC meets these three criteria. We can say that MiD, which has developed from EIC, continues to meet these criteria and succeeds in being a formal functional explanation (cf. Hawkins (2004: 13, fn. 1)). Hawkins argues that MiD offers a potential explanation for adjacency in grammatical rules and performance, e.g., multiple processing relations favor the adjacency of heads and complements more than between heads and adjuncts. (I will look at some examples of adjacency effects in 3.1.2.)

The principle of MiF follows from the same logic as MiD: if formal units and their associated syntactic and semantic properties are preferably reduced in the domains that must be accessed for combinatorial and dependency relations, then why not reduce the individual linguistic forms themselves, i.e. phonemes, morphemes, words, and phrasal units?

The principle of MaOP concerns efficiency differences between alternative ordering of relevant constituents, and seems to reduplicate some effects with MiD. A difference may be that MaOP is more dynamic than MiD (or EIC) in that MaOP involves efficiency in “on-line property assignments.” The basic intuition that the principle tries to capture is that many preferences appear to be correlated with the earlier assignment of common properties in one ordering vs. their later assignment in another.

2.2. Testing the Multiple Preferences of Efficiency Principles

Chapters 4–8 test the predictions of the Performance-Grammar Correspondence Hypothesis (PGCH, defined in (1)) on comparative data
from performance and grammars. The grammatical predictions of the PGCH are set out as in (6) (Hawkins (2004: 6)).

(6) Grammatical predictions of the PGCH

a. If there is a preference ranking A > B > C > D among structures of a common type in performance, then there will be a corresponding hierarchy of grammatical conventions (with cut-off points and declining frequencies of languages).

b. If a structure A is preferred over an A' of the same structural type in performance, then A will be more productively grammaticalized, in proportion to its degree of preference; if A and A' are more equally preferred, then A and A' will both be productive in grammars.

c. If two preferences P and P' are in (partial) opposition, then there will be variation in performance and grammars, with both P and P' being realized, each in proportion to its degree of motivation in a given language structure.

According to the PGCH, principles of performance should be reflected in the conventionalized rules of grammars and in grammatical variation. The predictions in (6) all refer to “performance preferences,” which are determined by the three efficiency principles in (3)-(5). Hawkins argues that there is a strong empirical basis to the PGCH, and performance preferences are supported by extensive data.

Let me briefly take the third case in (6c), where two opposing preferences are realized in grammars. An example is the symmetrical presence of the relative ordering of a verb (V) and a full NP direct object (O) across languages. Both V+O and O+V basic word orders are productive (see Hawkins (2004: 224) for figures), and in this case, each order is potentially motivated by either MaOP or MiD.4 For example, since V and O are mutually dependent, the V+O order guarantees backward-looking dependencies for some properties, but a delay in those property assignments for which V is dependent on O. The O+V order has the reverse advantages and disadvantages. Hence each ordering is motivated by MaOP, for at least some property assignments. In other

4 The MiF principle is irrelevant to this case.
words, a language may have either order from the performance point of view. Note, however, that processing efficiency tells this much, and whether a language chooses V+O or O+V is not a matter of performance (cf. Svenonius (2000: 7), but cf. Hawkins (2004: 11)). (I will come back to this in 4.1.)

2.3. More on the Performance-Grammar Correspondence Hypothesis

The general insights achieved in this book appear in the final chapter. Following the summary of the data supporting the PGCH in (1), Hawkins draws attention to grammatical generalizations that are either not predicted or heavily stipulated and for which the principles proposed in this book offer an explanation. In doing so, Hawkins does not blindly accuse generative grammarians (or more precisely, formalists) of neglecting the impact of performance factors on grammars. Instead, he grants that “generative grammarians are to be credited with the discovery of many [...] patterns” (p. 263) such as the distribution of gaps in easier positions and resumptive pronouns in more complex ones. He states that certain variants of Optimality Theory are one step ahead in gaining greater descriptive adequacy by incorporating performance preferences and hierarchies directly and seek some functional grounding and explanation for the constraints themselves (p. 264). Here, Hawkins stresses the importance of “formalization,” saying that “[f]ormalization can make explicit how functional generalizations have been conventionaled in grammars, and it can define all and only the predicted outputs of grammars (‘factorial typology’) with greater precision” (pp. 264–265). This statement represents Hawkins’ theoretical position as a formal functionalist. He goes further than Optimality Theory in seeking a theory for why the basic hierarchies and functional principles reflecting grammars are the way they are, i.e. “the ultimate explanation” for basic principles (p. 265). He formalizes the PGCH in (1) as the central guideline of the project, and tries to correct a view of the grammar-performance relationship that has been very widely accepted since Chomsky (1965), with the belief that “grammatical formalizations can and should incorporate the potentially explanatory principles derived from performance data, and psycholinguists should pay more attention to conventionalized grammatical variation than they currently do” (p. 267).
3. Issues Raised by a Formal Functionalist Approach

This book is the most recent general manifesto of Hawkins’ performance-based approach to grammars. Hawkins argues that there is a profound correspondence between performance and grammars (i.e. the PGCH in (1)), and shows that performance data can lead to new descriptive generalizations and to a new understanding of why grammars are the way they are. Although Hawkins (2004) largely depends on, and develops from, his earlier work (e.g. Hawkins (1990, 1994)), this book is motivated by new data and new insights. In this section, I will first point them out, and discuss some issues they raise regarding cross-linguistic variations in linearization (3.1). Then, I will consider specific problems in Hawkins’ implementation of the PGCH, namely, how to calculate “weight” which affects linearization, and then propose an alternative prosodically based definition of weight. I will also compare Hawkins’ competition based evaluation of weight with other competition theories, and point out potential problems with Hawkins’ system (3.2).

3.1. Going Beyond Syntax

Compared to Hawkins (1994) that appeared a decade earlier, Hawkins (2004) contains data from a smaller number of languages, but more thorough and systematic analyses of each linguistic phenomenon. Given this, it is not possible or sensible to tell which book covers wider empirical domain than the other. Among those differences between the two books, the most noticeable development is that Hawkins (2004) tries to apply his performance-based principles not only to syntax but also to morphology, morphosyntax, and semantics (cf. Hawkins (2004: 9)). Let us look at MiF first (3.1.1), and then MiD and MaOP (3.1.2) to illustrate this point.

3.1.1. Minimize Morphological Forms

The MiF principle in (4), in particular, makes predictions for the morphology of languages, because a linguistic form F includes not only words or phrases, but also morphemes. The minimization preference applies to the set of formal units that comprise a given F. For example, a long word, television, can be reduced to a shorter form, TV. Number markings can be expressed by an explicit morpheme (singular marking on nouns in Latvian) or by zero (English singulars). The MiF
principle defines an increasing preference for minimality in formal marking, ultimately for zero.⁵

(7) Reduce the formal units in \{F\} (Hawkins (2004: 39))
\[ F_0: P_1 > F_a: P_1 > F_{ab}: P_1 > F_{abc}: P_1 \]
(The “>” sign stands for preference. Small letters stand for arbitrary formal units of a given form F signaling a property P, with \( F_0 \) standing for zero formal marking.)

Note that the preference for minimality in form is calculable only when given forms are signaling the same property \( P_1 \).

3.1.2. Early Immediate Constituents, Minimize Domains, and Beyond

The principles of MiD (repeated below as (8) with my italics) and MaOP in (5) calculate not only syntactic but also semantic properties.

(8) Minimize Domains (MiD)
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. [...] It is easily seen that MiD is much more general than EIC in Hawkins (1994) (cf. 2.1) that defined a preference for minimal Constituent Recognition Domain (CRD).

(9) Constituent Recognition Domain (CRD) (Hawkins (1994: 58–59))
The CRD for a phrasal mother node M consists of the smallest set of terminal and non-terminal nodes that must be parsed in order to recognize M and all I(mmediate) C(onstituents) of M.

(10) Early Immediate Constituents (EIC) (Hawkins (1994: 77))
The human parser prefers linear orders that maximize the IC-to-non-IC ratios of CRDs.

The CRD and the EIC are defined purely in terms of syntax and syntactic processing (i.e. parsing).

The issue of whether the more general principle (MiD) is better than the less general one (EIC) should be both theoretical and empirical.

⁵ Although MiF is new in the context of PGCH, the idea builds on a large literature involving Haiman’s principle of economy, Levinson’s minimization principle, and so on (Hawkins (2004: 38)).
Theoretically, the more general the principle is, the more explanatorily adequate it should be. By taking “semantic” properties into consideration, however, one might wonder whether MiD and MaOP may be susceptible to lowered falsifiability, which leads to an empirical question. Let me take an example of adjacency effects in English and Japanese (cf. Hawkins (2004: ch. 5)), and show how MiD can be better than EIC. The case in point is the alternation between [V NP PP] and [V PP NP] in English (Heavy NP Shift), that between [V PP1 PP2] and [V PP2 PP1] in English (PP Shift), and that between [PP NP V] and [NP PP V] in Japanese (Scrambling).

For English Heavy NP Shift, the basic idea is that relative heaviness of the NP motivates the selection of [V PP NP] over [V NP PP] and this much is what EIC predicts (cf. Hawkins (1994: 182–185)). However, Wasow (1997) has shown that it is not just relative weight of verbal dependents, but also the semantic relation between V and PP that determine word order (see also Farrell (2005)). The more opaque the lexical combination between V and PP is (e.g. take into account X), the more adjacent (i.e. [V PP NP]) they become. Heavy NP Shift can apparently minimize surface domains for the processing of lexical combinations and dependencies within what Hawkins calls Lexical Domain (LD), the smallest domain sufficient for the processing of lexically listed combinations and dependencies. Lexical dependencies include semantic as well as syntactic dependencies, and MiD, which calculates not only relative weight but also lexical semantic relations involved in the alternation, predicts that lexical dependencies should all prefer adjacency. The table in (11) shows the numbers of [V PP PP] order, sorted out in terms of (i) weight and (ii) lexical dependency. In the table, Pd stands for the PP that is interdependent with V based on explicit verb entailment tests (see Hawkins (2004: 114, fn. 8) for details); and Pi stands for the PP that is independent of V by the same entailment tests.


<table>
<thead>
<tr>
<th></th>
<th>Weight</th>
<th>Lexical Dependency</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>Pd &gt; Pi by</td>
<td>Pd = Pi by</td>
</tr>
<tr>
<td></td>
<td>5+</td>
<td>2–4</td>
</tr>
<tr>
<td></td>
<td>7%</td>
<td>33%</td>
</tr>
<tr>
<td>[V Pd Pi]</td>
<td>8%</td>
<td>4%</td>
</tr>
</tbody>
</table>

(Proportion of adjacent V-Pd to non-adjacent orders given as a percentage.)
What do we see from the table (11)? Overall, multiple preferences with respect to (i) weight and (ii) lexical dependency have an additive adjacency effect by minimizing LDs. The interesting cases are when Pd is heavier than Pi by 1 word and when Pd is as heavy as Pi. The weight effect alone predicts that the [V Pi Pd] order is preferred in the former case, and that both orders are equally productive in the latter. However, the fact is that [V Pd Pi] is preferred over [V Pi Pd], resulting in adjacency between V and Pd, in both cases (74% in the former, and 83% in the latter). This is predicted by the second factor, lexical dependency: Pd wants to be adjacent to V more than Pi does. That is to say, it is multiple preferences by (i) and (ii) that determine the overall alternation pattern. Going back to the comparison between MiD and EIC, MiD correctly predicts the alternation pattern that EIC cannot predict and leaves as an exception.

Now let us turn to the mirror-image pattern in Japanese Scrambling. The alternation between [PP\textsubscript{m} NP\textsubscript{o} V] and [NP\textsubscript{o} PP\textsubscript{m} V] in Japanese (NP\textsubscript{o} = direct object NP with accusative case particle o; and PP\textsubscript{m} = PP constructed on its right periphery by a postposition) is comparable to the alternation between [V PPd PPi] and [V PPi PPd] in English, in that the verb contracts more syntactic and semantic relations with a direct object NP as a second argument or complement than it does with a PP, many or most of which will be adjuncts rather than complements (p. 118).

(12) Weight and direct object adjacency in Japanese

<table>
<thead>
<tr>
<th></th>
<th>n = 244</th>
<th>NP\textsubscript{o} &gt; PP\textsubscript{m} by</th>
<th>NP\textsubscript{o} = PP\textsubscript{m}</th>
<th>PP\textsubscript{m} &gt; NP\textsubscript{o} by</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5+</td>
<td>3-4</td>
<td>1-2</td>
<td>1-2</td>
</tr>
<tr>
<td>[PP\textsubscript{m} NP\textsubscript{o} V]</td>
<td>21%</td>
<td>50%</td>
<td>62%</td>
<td>66%</td>
</tr>
<tr>
<td>[NP\textsubscript{o} PP\textsubscript{m} V]</td>
<td>79%</td>
<td>50%</td>
<td>38%</td>
<td>34%</td>
</tr>
</tbody>
</table>

(Proportion of adjacent NP-V to non-adjacent orders given as a percentage.)

The overall pattern in Japanese Scrambling duplicates that of English PP Shift: when weight and lexical dependency reinforce each other in favor of [PP\textsubscript{m} NP\textsubscript{o} V] as in the right-hand columns, the result is significantly higher NP adjacency (of 80%, 84%, and 100%); when weights are equal there is a strong (66%) NP adjacency preference defined by lexical dependency preference alone; and when weight and lexical dependency are opposed as in the left-hand columns, the results are
split and the weight preference applies in proportion to its degree of preference.

To sum up, MiD correctly predicts the adjacency effects in English and Japanese, which are the results of multiple preferences: relatively heavy elements prefer to be non-adjacent to V, and lexically dependent elements prefer to be adjacent to V, minimizing relevant domains.

Let me take one further step. A simple comparison between the table (11) for English PP Shift and the table (12) for Japanese Scrambling suggests that lexical dependency effect may be stronger in English than in Japanese. For example, when the two phrases are of the same weight, adjacency between V and Pd is realized in 83% cases in English, whereas adjacency between V and NP0 is realized in 66% cases in Japanese. The difference suggests that the alternation between the two orders may be freer in Japanese (although a statistical analysis should be run to see whether this difference is significant or not). This is compatible with previous observations or analyses regarding Japanese (VP-internal) Scrambling, e.g. that neither order is more marked than the other and both orders can be base-generated (e.g. Miyagawa (1997) and Shiobara (2004), but see e.g. Hoji (1985), Takano (1998), and Yatsushiro (2003) for opposite views). Moreover, although some performance data suggests that the long-before-short weight effect is observed in Japanese (Hawkins (1994), Yamashita and Chang (2001)), such a preference is not found in acceptability judgments by native Japanese speakers (Tokizaki (2001), Shiobara (2004)).

My speculation at this point is that the relatively larger freedom in word order alternation in Japanese compared to English has to do with other interacting factors such as the availability of morphologically overt case particles, and head-finality. For example, it is reasonable to think that the case particle o on NP0 makes the NP semantically less dependent on the V, because we can tell at least that the NP is a direct object of the upcoming V before accessing the V. If so, the NP does not necessarily prefer to be adjacent to V. Whether a language has a morphologically overt case particle (e.g. Japanese) or not (e.g. English) is a matter of grammars, and does not seem to derive from any functional considerations. In this sense, the cross-linguistic linearization

---

6 Hawkins (2006) introduces another processing principle “Arguments precede X (ArgX)” to account for the relevant differences between VO and OV languages.
patterns are not determined solely by efficiency principles. I will come back to the interaction between efficiency principles and grammatical principles in section 4. (See also Neeleman and Reinhart (1998) and Takano (1998) for related discussion on Dutch Scrambling and Japanese Scrambling, respectively, and Kiaer and Kempson (2006) for relevant arguments on Korean Scrambling.)

With respect to the long-before-short weight effect in Japanese, such an effect is reported specifically for "production" data, e.g. text counts in Hawkins (1994) shown in the table (12) above, and the production experiment in Yamashita and Chang (2001), and not for comprehension such as acceptability judgments (Shiobara (2004)). I suspect that this has to do with the V-finality of Japanese: in production, V is anticipated by the speaker and hence lexical dependency information between V and its dependents is relevant, whereas in comprehension, dependents appear before their associated V and lexical dependency becomes relevant only after the V appears. If this speculation goes through, it will be the case that the weight effect interacts with a grammatical principle such as V-finality. I leave further investigation on this topic for the future (but see 4.1 for relevant discussion).  

3.2. How to Incorporate Weight and Where
3.2.1. How to Define Weight

The effects of weight on word order (whether alternative orders in performance or conventionalized order in grammars) constitute supporting evidence for the PGCH, because the notion of "weight" could not have a purely grammatical basis. Ross' (1967) Complex NP Shift, a predecessor of so called Heavy NP Shift, dictates that an NP is shifted under the condition that the NP dominates S(entence) and hence is complex. That is, Ross defines the complexity (or heaviness) syntactically. However, it is known that an NP that does not dominate S can be shifted. See (13). (See Hawkins (1990, 1994), Wasow (1997, 2002), Shiobara (2001, 2004), Akasaka and Tateishi (2003) among others for more examples.)

(13) a. Kay donated to the library five hundred Canadian dollars

Another factor worth investigating is prosody. See e.g. Neeleman and Reinhart (1998), Donati and Nespor (2003), and Shiobara (2004) for prosodic approaches to word order regularities within VP.
and her collection of novels by Mishima.

b. Ken ate in five minutes six BC rolls, four Alaska rolls, two or three dynamite rolls, and more than ten California rolls.

c. Ken ate in five minutes TEN CALIFORNIA rolls.

(Upper case letters indicate prosodic prominence.)

What contributes to heaviness of NPs in these examples? Hawkins' MiD, which subsumes EIC, calculates the weight by counting the number of words or terminal elements. This is empirically more adequate than Ross' definition of complexity, because it seems to explain cases like (13a, b) where the shifted NP contains relatively large number of words but not any Ss. However, the definition of "word" is not provided in Hawkins (1994, 2004), and we can only infer that he assumes something like a dictionary entry notion of wordhood. (See also Vicente (2005) for this point.) The notion of word is never trivial (cf. di Sciullo and Williams (1987)), and any argument based on word counting becomes unfalsifiable without defining what a word is. For example, one cannot know how Hawkins' word counting applies in polysynthetic languages such as Mohawk (Baker (1996)). Moreover, a simple word counting does not predict the well-formedness of examples like (13c), where the shifted NP (TEN CALIFORNIA rolls) consists of the same number of "words" (in Hawkins' sense) as the preceding PP (in five minutes).

Given these problems, I would like to suggest that weight should be counted in terms of the number of "prosodic words." In (13c), then, the shifted NP, made up of three prosodic words (TEN CALIFORNIA rolls), is heavier than the preceding PP, made up of two prosodic words with the stressless preposition (in) cliticizing onto the following lexical word (five). (See Zec and Inkelas (1990), Shiobara (2000, 2001), Akasaka and Tateishi (2003) for various prosodic definitions of weight.) The prosodically based definition of weight is more appropriate than a syntactically based one in light of processing because there seems to be evidence that sentence processing proceeds by prosodically defined "chunks" (cf. Gee and Grosjean (1983), Abney (1991), Shiobara (2000, 2001)).

3.2.2. How to Evaluate Weight

A rather theoretical issue arising from Hawkins' evaluation of weight is in its global counting system. First of all, MiD explicitly counts the
number of words (or whatever Hawkins assumes to be words), but this kind of counting does not seem to exist in other parts of the grammar. For example, there is no condition that makes reference to the checking of exactly three features, or to comparing the number of violations of some condition against the number of violations of some other condition, or whatever (cf. Collins (1997: 5)).

Regarding the problem of global comparison that MiD and MaOP necessarily make, let us recall the definitions of MiD and MaOP we saw in (3) and (5), respectively. Notice that the last sentences in the definitions in (3) and (5) include the notion of “competition.” According to MiD and MaOP, the human processor compares between linguistic sequences or structures and then picks out the optimal one. For example, among the two alternative orders [V PP1 PP2] and [V PP2 PP1] in English (e.g. John went in the late afternoon to London and John went to London in the late afternoon), the principles take into consideration all the relations of combination and/or dependency and properties assignable on-line to determine which order is more preferable than the other. When there are three PPs (e.g. John went [in the late afternoon] [to London] [after a long siesta]), it should be the case that six logically possible orders need to be considered to determine the optimal one. Impressionistically at least, this seems to require lots of computation and makes us wonder if the implementation of efficiency principles (e.g. MiD and MaOP) is efficient at all.

Before continuing on, it is worth noting that the issue of comparison is not new. For example, the computational component couched in the Minimalist framework (Chomsky (1995)) defines the “reference set” to determine the most economical derivation. Chomsky (1995: 227) notes that “selection of an optimal derivation in the reference set [...] poses problems of computational complexity too vast to be realistic” and the Minimalist approach has been moving into the direction of reducing the problem with a more local interpretation of reference sets, one of which became realized as the notion of “Phase” as a computational cycle (Chomsky (1998, 1999) and subsequent works, see Collins (1997) and Reinhart (2006) for discussion).

However, the problem of global comparison may have different fla-

8 An Optimality Theoretic model of grammar may be an exception to this.
vors in different models of grammar. In the Minimalist Program, which is guided in part by the question of what conditions are imposed on the language faculty by virtue of “general considerations of conceptual naturalness that have some independent plausibility, namely, simplicity, economy, symmetry, nonredundancy, and the like” (Chomsky (1995:1)), to reduce computational complexity is not an empirical necessity but still considered illuminating for deepening our understanding of the grammar. In contrast, Hawkins’ formal functionalist approach to the grammar is driven by the PGCH in (1), and hence argues that the grammar is shaped in response to performance efficiency, without any unnecessary computational complexity. Therefore, the real problem of global comparison incorporated in MiD and MaOP seems to boil down to its incompatibility with performance efficiency.

4. From Formal Functionalist to Functional Formalist

Let us finally turn to more general theoretical issues raised by Hawkins’ formal functionalist approach to the grammar. First, I consider the architecture of the language faculty in terms of how grammatical and performance principles interact to determine linear order (4.1). Then, I propose a functional formalist approach, which is instantiated as the left-to-right computation in the grammar (4.2). The section concludes with some final remarks (4.3).

4.1. Architecture of Language Faculty: Interaction of Principles

In implementing the PGCH, Hawkins (2004) proposes three efficiency principles, MiD, MiF, and MaOP, and auxiliary sub-principles (see 2.1). Since these principles are formulated from a viewpoint of the human processor, with all the definitions starting with the human processor prefers to..., one may be still unclear on how Hawkins’ formal functionalist approach is different from purely functionalist approaches. In order to clarify this point, let us consider how these performance-based principles interact to make collective predictions, not only for performance, but also for grammars.

For people who believe in PGCH, even highly abstract and fundamental properties of syntax may be argued to be derivable from the interaction of these externally motivated principles that are needed anyway in order to explain how language is used. Under this view, both cross-linguistic commonalities and variations are traced to how gram-
mars conventionalize/grammaticalize these performance-based principles, and to what extent. Let us take the position of adverbs, for example. There is a cross-linguistic tendency that adjuncts only intervene between a head and a complement if they are short and can minimize the relevant phrasal combination domain, thereby speeding up phrase structure processing (cf. Hawkins (2004: 133)). In the corpus of Hawkins (2000), single-word adverbs do regularly intervene between V and longer PPs in English, and there are languages in which verb-object bonding has been less strongly “grammaticalized” than in English, such as French, and in which short adjuncts frequently intervene, as in j’admire souvent la gloire de mon père (‘I admire often the glory of my father’).

The multiple principles make collective predictions. To take another example, recall the ordering [(V) NP PP] in English and its counterpart in Japanese we saw in 3.1.2. The overall ordering patterns in English and Japanese are predicted by the interaction of weight and lexical dependency. However, recall also that a difference was that the lexical dependency effect seemed weaker in Japanese than in English, resulting in relatively freer ordering in Japanese.

Having seen this, what seems missing in this book, or beyond the scope of this book, is discussion on the interaction between these performance-based principles and purely grammatical principles. Going back to the ordering examples again, Hawkins does not note, or provide explanation for, the difference in relative freedom in ordering between English and Japanese (but see note 6). In 3.1.2, I suggested the possibility that the freer ordering in Japanese might be due to independent grammatical properties such as head-finality and the existence of morphologically overt case particles. That is to say, the interaction of Hawkins’ performance-based principles is in charge of accounting for the common ordering pattern in English and Japanese, and my speculation is that the interaction of these principles with other grammatical principles is in charge of accounting for the difference in the ordering patterns between English and Japanese.

Another related issue concerns the relative strength of performance-based principles and grammatical principles. Vicente (2005) points out that there are several fixed expressions that forbid word orders that would result in a shorter processing domain, such as the (b) examples in (14)–(16).

(14)  a. give Mary the sack
b. *give the sack Mary
(15) a. throw John to the lions
b. *throw (to) the lions John
(16) a. drive Peter bananas
b. *drive bananas Peter

In (b), the idiomatic parts of the VP are brought together just like in (a). Despite no loss or gain in processing efficiency, however, the (b) sentences are totally ruled out. Vicente (2005) speculates that the grammar can generate a number of linguistic expressions which are more or less difficult for performance, but the reverse does not seem to be true: “[performance] considerations cannot force a structure that is not independently generable by grammar.” He goes on saying that “performance [principles] can select the most parser-friendly structure amongst a number of alternatives, but they cannot generate the best possible structure independently of syntax.”

Strictly speaking, however, the (b) examples in (14)–(16) do not argue against Hawkins’ performance-based approach to grammars, because Hawkins does maintain the distinction between grammar and performance systems, and never argues that every grammatical principle is performance driven. The PGCH only says that “grammars have conventionalized syntactic structures” and not that all the syntactic structures in grammars are conventionalized in response to preference in performance (see (1)). In other words, the PGCH is part of the project that seeks how far the performance-grammar correspondence goes, and it remains to be seen what is left over for the grammar to do on its own. Hawkins suspects not much. Therefore, the challenge that the (b) examples in (14)–(16) offer is their severe ill-formedness, which remains unpredicted by the MiD or the PGCH alone. These kinds of examples are worth further investigation for our deeper understanding of the performance-grammar interaction and the architecture of language faculty.

Let us look at another related example. The ordering of PPs of the same type (e.g. Pd or Pi, see 3.1.2) exhibits the increasing weight effect, as is predicted by MiD. That is to say, PPs are ordered from light to heavy in accordance with performance efficiency. For example, the (a) sentences are preferred over the corresponding (b) sentences in (17) and (18). (Examples are from Culicover and Jackendoff (2005: 146-147)).

(17) a. Robin talked with Dana about the cockroach that ate
(18) a. Robin talked about Leslie with the former vice-chairman of the sociopathy department.

b. ?Robin talked with the former vice-chairman of the sociopathy department about Leslie.

The question, however, arises as to whether any syntactic principles are needed for the case in which an NP is reordered with respect to PPs, as in (19).

(19) Fred discussed with the class [NP the strong constraints on long-distance wh-movement].

(Culicover and Jackendoff (2005: 147))

Compared to the alternation between two PPs, the ordering of NP and PP is much more restricted, and [NP PP] is much more frequent than [PP NP]. According to Hawkins’ (1994) text counts, the number of [NP PP] was 95% (458 out of 480), and [PP NP] was 5% (22 out of 480). This frequency effect suggests that [NP PP] should be the syntactic “default” order, and the order in (19) does not conform to it. This is the case where the relative strength of principles comes into play: when the requirement of MiD wins over the syntactic default, [PP NP] is achieved. (See Shiobara (2004, 2006) and Culicover and Jackendoff (2005: 147) for related discussion.) At this point, it is worth reminding ourselves that Hawkins (1994) gives performance-based reasoning even for the syntactic default order: since PP is, in principle, heavier than NP by having a preposition in addition to its complement NP, the default order [NP PP] is a reflection of the light-heavy ordering and hence the grammaticalized manifestation of the increasing weight effect. (But see 3.2.1 for the problem in counting in function words such as prepositions.) Likewise, Hawkins argues that a clause (CP or IP) appears after other dependents (e.g. NPs or PPs) in English, because a clause is potentially heavier than any other phrase by having a clausal layer, which is grammaticalized as fixed word orders. The same reasoning applies to German, where a CP complement always appears postverbally whereas an NP complement always appears preverbally (cf. Hawkins (1994: 293–308)).

If this performance-based account of ordering patterns is valid, we would expect that a relatively heavy NP may appear after a clause in English, which does not seem to be borne out.
Likewise, the ordering effect in German is categorical and does not seem to be influenced by the weight of CPs or NPs (cf. Inaba (to appear) and references therein). Based on the grammaticalization theory, Hawkins (or more generally, functionalists) might say that once grammaticalized, performance-based principles do not show their effects any more. If so, however, whether a grammatical principle is of a purely grammatical origin or from a performance-based grammaticalized principle cannot be determined or tested. Let me summarize word order patterns we have seen so far which seem to be fixed as grammatical conventions.

21. a. V-initiality (V+O) and V-finality (O+V) across languages
b. V—XP—YP, but not V—YP—XP in English
   (e.g. XP = NP, YP = CP or IP, as opposed to relatively free alternation between PP—NP—V and NP—PP—V in Japanese, cf. section 3.1.2)
c. V—CP and NP—V in German
d. V—Adv(short adjunct)—O in French

We saw that Hawkins' formal functionalist approach to grammars may give potentially functional explanations for these word order patterns. For (21a), MaOP predicts that both orders should be productive across languages. However, it remains unexplained why some languages choose V+O and others O+V (see section 2.2). For (21b), a PP is in principle heavier than an NP and this weight effect is grammaticalized and realized in the fixed NP—PP order in English. Likewise, for (21c), a CP is in principle heavier than an NP and this weight effect is grammaticalized and realized in the fixed word order in German. For (21d), MiD predicts that only a short adjunct may intervene between verb and object, and how regularly this happens is subject to the degree of grammaticalization of verb-object bonding.

I argue that the word order patterns in (21) should not be given grammaticalization-based explanations because they are either wrong or impossible to test; rather, these patterns should be regarded as purely grammatical rules. First of all, as for (21a), processing efficiency or any other performance consideration does not tell why English is head-initial and Japanese is head-final. Head-directionality is just a matter of how their grammars are. (Recall that performance efficiency does
not predict the conventionalized word order patterns we saw in (14)–(16) either.) Secondly, as I briefly noted above (21), appealing to
degree of grammaticalization or grammaticalization itself makes
Hawkins’ formal functionalist approach less formal and hence not
testable. In order to get the grammaticalization scenario to work, we
need independent evidence for that such as diachronic change in word
order, e.g., English used to allow the V—PP—NP order more freely
than now when NP was relatively heavier than PP. Without such inde-
pendent evidence, I would argue that these ordering patterns are just the
way grammars are and solely based on grammatical factors. As for
(21b–d), I speculate that the cross-linguistic differences as to the free-
dom of rearrangement of dependents are due to independently observed
grammatical factors, such as head-directionality, availability of morpho-
logically overt case particles (see 3.1.2), presence or absence of mor-
phological agreement between verb and object, and so on.

4.2. Left-to-Right Computation in the Grammar

Having granted the presence of purely grammatical rules or principles
that are not overridden by performance factors, I am led to depart from
the (formal) functionalist position such as Hawkins’, moving into the
formalist position. Look at the relevant cells of the table (2) with
additional information below.

(2’) Formalist and functionalist approaches to linearization

<table>
<thead>
<tr>
<th>Competence (grammar)</th>
<th>Relation</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>ii. Functional formalist (this review article)</td>
<td>L-to-R computation rules for (21a–d)</td>
<td>↔</td>
</tr>
<tr>
<td>iv. Formal functionalist (Hawkins (1994, 2004))</td>
<td>grammaticalized efficiency principles rule for (21a)?</td>
<td>↔</td>
</tr>
</tbody>
</table>

In Hawkins’ formal functionalist approach (iv), efficiency principles,
which are independently needed for explaining performance data, are
largely grammaticalized and shape the grammars. It is not clear how
many of the grammatical principles are purely grammatical, but whether
a language is head-initial or head-final (e.g. (21a)), for example, should
be one of the purely grammatical rules that are used in performance.
My functional formalist position (ii) is motivated by the existence of purely grammatical rules of linearization such as those for (21a−d) which are not overridden by performance factors. This is where my position differs from Hawkins'. On the other hand, I agree with Hawkins that the relation between the grammar and performance systems is tighter than assumed by pure formalists. In particular, I propose that the grammatical computation proceeds from left to right, as opposed to from bottom up, in the same way as on-line processing in performance does (cf. Shiobara (2004), following Phillips (1996, 2003), see (22) below for illustration). Note, however, that I do not assume a cause-effect relation between grammar and performance. It is only that the grammar trivially meets the PF interface condition that terminals are linearized from left to right by using the same left-to-right (L-to-R) derivation and syntax-PF mapping.

In addition to L-to-R derivation, I adopt Chomsky’s (1998, 1999) Phase-based computation, and propose that syntactic objects are spelled out as prosodic objects, namely as intonational units (Φ). For example, the English sentence *Ken saw Naomi* and its Japanese equivalent *Ken-wa Naomi-o mita* (*Ken-Top Naomi-Acc saw*) are derived as in (22): syntactic structure is built from left to right and spelled out as a prosodic object iteratively.

\[
\begin{align*}
\text{(22) a. English} & \quad \text{b. Japanese} \\
\text{Syntax: (merge from left to right)} & \quad \text{PF: (spell out from left to right by Φs)} \\
\text{Ken saw } & \quad [Φ \text{ Ken-wa}] \quad \rightarrow [Φ \text{ Naomi-o mita}]. \\
\text{Ken} & \quad \rightarrow [Φ \text{ Ken-wa}] \\
\text{saw Naomi} & \quad [Φ \text{ Naomi-o mita}] \\
\end{align*}
\]

This proposal is inspired by Phillips’ (1996, 2003) work, in which he provides independent syntactic evidence for L-to-R syntactic derivation. In this model, the grammatical rules on linearization such as those we saw in (21) are used within syntactic derivation. For example, English merges the subject with the verb first, not with the object, because English is V-initial. On the other hand, Japanese merges the subject with the object, not with the verb, because Japanese is V-final.

The next step is to show how L-to-R computation in grammars can
account for performance data. Remember that one of the strengths of Hawkins' formal functionalist approach is that performance efficiency accounts for not only performance, but also (typological) grammatical variation. In order to show the strength of the functional formalist approach, then, we need to look for the case where L-to-R derivation in the grammar may be appropriate to account for certain performance variation. I argue that the weight effect in English Heavy NP Shift is the case in point: it should be prosodically captured (see 3.2), and hence be naturally regarded as a PF interface condition. It has been noticed that the NP in the V—PP—NP order in English forms its own intonational domain (Φ) (Rochemont and Culicover (1990), Zec and Inkelas (1990), Zubizarreta (1998)), and hence we can formulate a PF interface condition that “the prosodic object F which is spelled out finally must contain more prosodic words than the F which is spelled out second finally.” This is to say, syntax can generate both V—NP—PP and V—PP—NP orders, and it is the PF interface condition that chooses one over the other. This way, the weight effect is purely prosodically defined, and the weight evaluation is limited to local, adjacent Φs (= Phases), resolving the problem of global computation (at least partially, cf. 3.2.2). I basically assume here, without giving thorough arguments due to space limitation, that what is spelled out is no more accessible to the computation at the next Phase level, yielding the effect of the Phase Impenetrability Condition (Chomsky (1998, 1999, 2004)).

4.3. Epilogue: Going Beyond Explanation

The main argument in this book is that there is a profound correspondence between performance and grammars, and that common preferences of performance and grammars can be explained by efficiency and complexity. This argument can only be convincing because it is given by a formal, functional linguist such as Hawkins, who has wide and in-depth knowledge of performance and grammatical data, as well as functional and formal approaches. The performance data come from corpus studies and processing experiments, and the grammatical data come from typological samples and the growing number of languages that have now been subjected to in-depth formal analysis. Probably due to his well known comparative studies of languages as a typological linguist (e.g. Hawkins (1983, 1988), to name just a few), Hawkins’ work may be considered more descriptive than explanatory. This book clear-
ly corrects this one-sided view, and succeeds in reconciling the results of formal and functional approaches to the language faculty to a great extent.

For those that are familiar with formal theories of grammars, it may not be difficult to connect Hawkins’ thirst for the “ultimate explanation” for grammatical principles (Hawkins (2004: 265), cf. section 1) with the Chomskyian program of moving to a deeper level of explanation, “beyond explanatory adequacy” (Chomsky (2004, 2005, to appear)). Both deal with the question of why grammars are the way they are, seeking for a principled explanation of grammars. Furthermore, Chomsky mentions concepts of “computational efficiency that carry us beyond explanatory adequacy” (Chomsky (to appear), italics are mine). This indicates that Chomskyan formalist position shares with Hawkins’ formal functionalist position certain methodological considerations to adhere to generative systems with computational efficiency and without excessive redundancy or unattractive properties such as complexity. Their actual approaches appear to be different in that Chomsky attempts to crystallize efficient grammar from inside, i.e. by investigating grammatical data, whereas Hawkins attempts to do so not only from inside, but also from outside, i.e. by investigating not only grammatical but also performance data. Furthermore, while Hawkins refers to “computational efficiency” as a notion specific to language performance, Chomsky takes it as a general, language-independent notion (as in e.g. Chomsky (2005: 10)). Despite these differences in their approaches, Chomskyan formalist and Hawkins’ formal functionalist approaches have common concerns about why grammars are the way they are. In this review article, I took a functional formalist position in an attempt to extract good parts from these two related approaches.

Let me end this review with a brief and rather personal recommendation. To me, one of the most impressive aspects of this book is the consistently modest attitude of the author. Hawkins says at the outset that “I do not currently know how far the PGCH can be taken and to what extent it can replace proposals for an innate grammar” (Hawkins (2004: 11)). He goes on to say that “[w]e need to keep an open mind on this whole issue and avoid dogmatic commitments to uncertainty.” To me, it is his open-mindedness as a linguist of curiosity, on top of his thorough knowledge of languages and linguistic theories, that make this book a landmark which not only opens up new territory for discussion, but also bridges the gap between formalist and functionalist
approaches to grammars.

REFERENCES


Department of Foreign Languages
Bunkyo Gakuin University
1-19-1 Mukougaoka, Bunkyo-Ku
Tokyo 113–8668
email: kayono@fs.u-bunkyo.ac.jp