Problems in the Language Form and Function Debate and a Way Forward

John A. Hawkins
University of California, Davis, USA & Cambridge University, UK

Abstract

A much debated issue in linguistics involves the extent to which the forms of languages (phonemes, morphemes, words, phrases and syntactic structures) have been shaped by the functions that languages perform. This paper begins by defining what these "functions" are: the forms of language must express (i.e. be capable of being paired with) an infinite set of meanings in a consistent way; these form-meaning pairings must be readily processable by language producers and comprehenders in real time; and they must be readily learnable. This paper then focuses on the processability function, and enumerates some problems that complicate discussions of whether processing has actually played a role in shaping grammars, especially in morphology and syntax. Most of these problems, it is argued, are the result of unresolved issues in current theories regarding e.g. the precise relationship between production and comprehension, the measurement of working memory load, and the relationship between prediction and the integration of preceding and following items in on-line processing. I argue here that the linguistic question of explaining why the world's grammars are the way they are, and understanding the role of processing in explaining grammatical forms, need not, and cannot, wait for these big general issues to be resolved in psycholinguistics. Instead, we can adopt a more empirical approach that compares patterns in grammatical forms and structures across languages with relevant empirical patterns in usage data within languages, from corpora and experiments. This more empirical approach allows us to test whether there is, or is not, a match between grammatical data on the one hand, and usage data on the other, that are the products of the processing mechanisms that psycholinguistic theories are trying to define. These patterns can be described in theory-general or -neutral ways and they can shed light on whether grammatical forms have been shaped by performance, without us having to fill in all the details of the contributing processing theories. This paper illustrates two empirically based descriptive principles and their supporting data, Minimize Domains and Minimize Forms, and considers these methodological issues.

Keywords: efficiency, form-function correspondences, functions of language, head ordering, hierarchies, language processing, Minimize Domains, Minimize Forms, Performance-Grammar Correspondence Hypothesis

1. Forms and Functions

When linguists debate the relationship between form and function in language (as in Frederick Newmeyer's influential 1998 book) they are essentially addressing two questions. First, to what extent have the forms of languages (phonemes, morphemes, words, phrases and syntactic structures) been shaped by the functions that languages perform? Second, what exactly are these "functions"?

With respect to the second question there are three very general functions that languages can be said to perform: (i) the forms of language must express (i.e. must be capable of being paired with) an infinite set of meanings in a consistent way; (ii) these form-meaning pairings must be readily processable by language producers and comprehenders in real time; and (iii) they must be readily learnable by children.

Everyone agrees with regard to (i) that all human languages are capable of expressing an infinite set of meanings. Disagreements tend to center on more technical matters of how closely the rules of syntax and semantics correspond, at what levels of abstraction semantic interpretation takes place, the interplay between semantic interpretation and pragmatic enrichment for formal entities, and so on. For (ii) processing and (iii) learning the disagreements are more profound. Clearly all languages are readily processable in real time, but do processing and performance have any role to play in actually explaining the formal properties of language(s)? In other words, does processing have a productive role to play in the explanation of formal grammars? Some think not. Likewise, all languages are readily learnable, but what role, if any, has learning played in actually explaining and shaping the formal properties of the languages being learned?

In this paper I focus on the processing function in relation to the forms of languages, especially morphology and syntax. I define in section 2 what I see as the problems in the form-function debate in this area, and I then outline in sections 3-5 an empirically based research methodology that can contribute to their resolution. Section 6 summarizes my conclusions.

2. The Problems

Most of the problems that hinder our ability to achieve clarity and agreement on the form-function relationship stem, I would argue, from current unclarities and unresolved issues in the contributing psycholinguistic areas of processing and learning, combined with some traditional and entrenched views by linguists. When there is unclarity over what some basic function is that is claimed to explain certain forms, then this makes it difficult to determine whether that function does indeed underlie those forms. And if some group of linguists holds the fixed belief that certain forms are not explainable through certain

functions, but are autonomous and perhaps innate, then no amount of empirical data is going to change their minds.

Some big psycholinguistic issues that require more clarification and research at the moment and that are often discussed in the context of functional explanations for grammars include the following (with sample publications in parentheses that the reader can consult for full elaboration of the issues):

(a) the relationship between production (the speaker) and comprehension (the hearer) (Wasow, 2002; Hawkins, 2014, pp. 50-51);
(b) the nature of working memory load and capacity constraints (Just & Carpenter, 1992; Gibson, 1998, 2000; Hawkins, 2014, pp. 47-49, 232-233);
(c) the role of integration vs. prediction in on-line processing (Vassisth, 2011; Levy, 2008; Hawkins, 2014, pp. 59-61, 233-234);
(d) the role of frequency and its relationship to other determinants of ease of processing (Bybee, 2007, 2010; Hawkins, 2004, pp. 41-49, 64-68);
(e) the precise definition of grammatical complexity (Newmeyer & Preston, 2014; Hawkins, 2009);
(f) the relationship between efficiency and ease of processing (Jaeger & Tily, 2011; Hawkins, 2014, pp. 47-49, 230-232);
(g) the relative strengths of competing and cooperating principles (MacWhinney, Malchukov, & Moravcsik, 2014; Hawkins, 2014, ch. 9);
(h) some traditional views in linguistics to the effect that performance has not shaped grammars (Chomsky, 1965; see Newmeyer, 1998, Hawkins, 2014, pp. 62-72, 225-230 for discussion);
(i) the relationship between ease of processing and ease of learning (O'Grady, 2005; Filipovic & Hawkins, 2013).

For example, with respect to (a) it is not currently clear to what extent the details of a production model (as in e.g. Levelt, 1989) incorporate the same or similar mechanisms to comprehension, as reflected in the much larger experimental findings on parsing and comprehension (see Traxler, 2011 for an overview). This issue is particularly acute for the production of head-final languages like Japanese, that have mirror image orderings of categories to head-initial languages like English, and when the verb and other head categories are final in their respective phrases in contrast to the head-initial orders of English. Head-final orders are potentially associated with very different on-line processing routines for producers versus comprehenders. The purist would argue that if the details of these models are not clear, we cannot begin to assess whether or not they explain the relevant forms and structures of grammars. I disagree.

3. A Way Forward

I will argue here that for those of us whose primary research goal is the linguistic one

of explaining why the world's grammars are the way they are we need not, and cannot, wait for these big general issues to be resolved in psycholinguistics. We can grant the desirability of further progress in the contributing fields of language learning and production and comprehension (and in our understanding of working memory load, integration and prediction, frequency effects, etc.), but we can still make substantial progress as follows.

First, we can adopt a more empirical approach that compares patterns in grammatical forms and structures across languages with relevant empirical patterns in usage within languages, especially from written and spoken corpora. We can also compare patterns of grammars with empirical patterns from language learning (first and second). An empirical approach allows us to test whether there is, or is not, a match between conventionalized data from grammars on the one hand, and usage data and learning data on the other that are the products of the processing and learning mechanisms that psycholinguistic theories are trying to define.

Second, such patterns can be described in ways that are theory-general or -neutral and they can shed light on whether grammatical forms have or have not been shaped by processing and learning functions, without us having to fill in all the details of the contributing processing, learning and grammatical models.

Third, when searching for empirical patterns we can focus on those insights that are currently clear and agreed upon, rather than those that are less clear. And we can use the empirical support for what is clear to shed light on what is less clear, both in grammars and in processing/learning.

This is in essence the research program that I have adopted and described in my 2004 and 2014 OUP books *Efficiency and Complexity in Grammars* and *Cross-Linguistic Variation and Efficiency*, and in principles such as Minimize Domains that are presented there and in earlier publications. These are essentially empirically based descriptive patterns that are claimed to hold for both cross-linguistic grammatical forms and for usage data. They are not, as some have suggested, principles of comprehension only, of no relevance for production. Minimize Domains is a general descriptive pattern drawn from both comprehension data and (even more extensively) production data, and from typological samples of grammatical data, i.e. from the three sources of grammatical and usage data for which we have empirical evidence. These data provide prima facie support for the claim that the processing function of language has indeed shaped grammatical form. This approach remains theory-neutral about how exactly the general descriptive patterns I have enumerated should be implemented in a production model, in a comprehension model, and in a formal model of grammar. It remains neutral as to how exactly the conventions of grammars emerge in historical time. These matters need more research. The empirical support for Minimize Domains in production, comprehension and grammars is clear and strong, however, and this empirical support can now shed light on the details of the respective models for each of these areas, and on historical mechanisms, that are currently less clear. The empirical correlations between production

data, comprehension data, and grammars are directly relevant for the primary research question raised at the beginning of this paper: has the processing function of language shaped the forms and structures of the world's grammars? The empirical evidence, I suggest, is an overwhelming yes. In what follows I briefly illustrate this methodology and the correlating empirical patterns between processing and grammars that it has uncovered.

4. Minimize Domains

In Hawkins (2004, 2014) I proposed three general efficiency principles, one of which is Minimize Domains.

This principle is evident in both performance and grammars and it involves the size of the syntactic domain in which a given grammatical relation can be processed. How great is the distance separating interrelated items and how much other material needs to be processed simultaneously as this relation is processed? In those languages and structures in which domain sizes can vary in performance, we see a clear preference for the smallest possible domains. In those languages and structures in which domain sizes have been grammatically fixed, we see the same preference in the conventions. The organizing principle here is defined in (1) (see Hawkins, 2004, p. 31):

(1) 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.

Consider the ordering of words. Words and phrases have to be assembled into the kinds of groupings that are represented by tree structure diagrams as they are parsed and produced in the linear string of speech. Assigning phrase structure can typically be accomplished on the basis of less than all the words dominated by each phrase. Some orderings reduce the number of words needed to recognize and construct a mother phrase M and its immediate constituent daughters (ICs), making phrasal combination faster. Compare (2a) and (b):

(2) a. The man vp[waited pp1[for his son] pp2[in the cold but not unpleasant wind]]

```
1          2   3   4          5
----------------------------------
```

b. The man vp[waited pp2[in the cold but not unpleasant wind] pp1[for his son]]

```
1          2  3    4    5   6        7          8          9
-------------------------------------------------------------------
```

The three items, V, PP_1, PP_2 can be recognized and constructed on the basis of five words in (2a), compared with nine in (2b), assuming that (head) categories such as P immediately project to mother nodes such as PP, enabling the parser to recognize and construct them on-line. Minimize Domains predicts that Phrasal Combination Domains (PCDs) (see the definition in (3)) should be as short as possible, and the degree of this preference should be proportional to the minimization difference between competing orderings. This principle was called Early Immediate Constituents (EIC) in Hawkins (1994), and defined in (4):

(3) 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 (plus all M-dominated non-terminals over the terminals) on the basis of which the processor can construct M and its ICs.

(4) Early Immediate Constituents (EIC) [Hawkins, 1994, pp. 69-83]
The human processor prefers linear orders that minimize PCDs (by maximizing their IC-to-word ratios), in proportion to the minimization difference between competing orders.

In concrete terms EIC amounts to a preference for short before long phrases in head-initial structures like those of English, e.g. for short before long PPs in (2). These orders will have higher "IC-to-word ratios", i.e. they will permit more ICs to be recognized on the basis of fewer words in the terminal string. The IC-to-word ratio for the VP in (2a) is 3/5 or 60% (5 words required for the recognition of 3 ICs). The comparable ratio for (2b) is 3/9 or 33% (9 words required for the same 3 ICs). (For comparable benefits within a Production Model, see Hawkins, 2004, p. 106).

Structures like (2) were selected from a corpus on the basis of a permutation test (Hawkins, 2000, 2001): the two PPs had to be permutable with truth-conditional equivalence (i.e. the speaker had a choice). Only 15% (58/394) of these English sequences had long before short. Among those with at least a one-word weight difference (excluding 71 with equal weight), 82% had short before long, and there was a gradual reduction in the long before short orders, the bigger the weight difference (PP_S = shorter PP, PP_L = longer PP):

(5) n = 323	PP_L > PP_S by 1 word	by 2-4	by 5-6	by 7+
[V PP_S PP_L]	60% (58)	86% (108)	94% (31)	99% (68)
[V PP_L PP_S]	40% (38)	14% (17)	6% (2)	1% (1)

For experimental support (in production and comprehension) for short before long effects in English, see e.g. Stallings (1998), Gibson (1998), Wasow (2002).

A possible explanation for the distribution in (5) can be given in terms of reduced processing demands in working memory. If, in (2a), the same phrase structure information can be derived from a 5-word viewing window rather than 9 words, then phrase structure processing can be accomplished sooner, there will be fewer additional (phonological,

morphological, syntactic and semantic) decisions that need to be made simultaneously with this one, and there will be less structural complexity to compute and fewer competing structural decisions to resolve (cf. Lewis & Vasishth, 2005). Overall fewer demands will be made on working memory and on the computational system. (2a) is more efficient, therefore, but not because some claimed capacity limit has been breached. All the attested orderings in (2) are clearly within whatever limit there is. The graded nature of these data point instead to a preference for reducing simultaneous processing demands when combining words into phrases. More generally we can hypothesize that the processing of all syntactic and semantic relations prefers minimal domains, which is what MiD predicts.

For head-final languages, long before short orders provide minimal domains for processing phrase structure. Consider (6):

(6) a. Mary ga [[kinoo John ga kekkonsi-ta to]s it-ta]vp
Mary SU yesterday John SU married that said,
'Mary said that John got married yesterday'
b. [kinoo John ga kekkonsi-ta to]s Mary ga [it-ta]vp

Why? Because when longer phrases are positioned before shorter ones in Japanese then constructing categories or heads (V, P, Comp, etc) are close, or as close as possible, to each other, each being on the right of their respective phrasal sisters. As a result Phrasal Combination Domains are smaller and more minimal.

Consider the following basic word orders of Japanese grammar in (7):

(7) a. Taroo ga vp[tegami o kaita] NP-V
T. SU letter DO wrote
'Taroo wrote a letter'
b. Taroo ga pp[Tokyo kara] ryokoosita NP-P
T. SU Tokyo from travelled
'Taroo travelled from Tokyo'
c. np[[Taroo no] ie] Gen-N
Taroo 's house

The heavier phrasal categories, e.g. NP, always occur to the left of their single-word heads in Japanese, e.g. before V and P, and P and V are then adjacent on the right of their respective sisters.

An early corpus study testing the long before short prediction for Japanese was given in Hawkins (1994) using data collected by Kaoru Horie for alternative orderings of [{NPo, PPm} V] structures such as (8):

(8) a. (Tanaka ga) [[Hanako kara]pp [sono hon o]np katta]vp
Tanaka SU Hanako from that book DO bought

'Tanako bought that book from Hanako'

b. (Tanaka ga) [[sono hon o]np [Hanako kara]pp katta]vp

The results are shown in (9) where IC_S = shorter Immediate Constituent; IC_L = longer Immediate Constituent; regardless of NP or PP status:

(9) IC_L>IC_S	by 1-2 words	by 3-4	by 5-8	by 9+
[IC_S IC_L V]	34% (30)	28% (8)	17% (4)	9% (1)
[IC_L IC_S V]	66% (59)	72% (21)	83% (20)	91% (10)

As can be seen in (9), the bigger the weight difference, the more the heavy phrase occurs to the left; i.e. this is the mirror-image of English. For experimental and corpus support for long before short phrases in Japanese and Korean when there is a plurality of phrases before V, see further Hawkins (1994, 2004), Yamashita and Chang (2001, 2006), Choi (2007).

Given these data from performance, supporting the processing efficiency principle Minimize Domains, we can now better understand:

(a) the Greenbergian word order correlations for word orders in grammars;
(b) why there are just two equally productive word order types across languages, head-initial and head-final; and
(c) why and when there are "exceptional" departures from the expected head-initial and head-final orders in grammars.

Grammatical conventions of ordering across languages reveal the same degrees of preference for minimal domains. The relative quantities of attested languages reflect the preferences. An efficiency approach can also explain exceptions to the majority patterns and to grammatical principles such as consistent head ordering. Full supporting data and discussion are given in Hawkins (1994, 2004, 2014). Consider some brief examples here.

Greenberg (1963) examined alternative verb positions across languages and their correlations with prepositions and postpositions in phrases corresponding to (10):

(10) vp{V, pp{P, NP}}

a. vp[**travels** pp[**to** the city]] b. [[the city **to**]pp **travels**]vp

c. vp[**travels** [the city **to**]pp] d. [pp[**to** the city] **travels**]vp

The language quantities for these different combinations in Dryer's (1992) cross-linguistic sample are shown in (11):

(11) a. vp[**V** pp[**P** NP]] = 161 (41%) b. [[NP **P**]pp **V**]vp = 204 (52%)

c. vp[**V** [NP **P**]pp] = 18 (5%) d. [pp[**P** NP] **V**]vp = 6 (2%)

The preferred (a)+(b) have consistent head ordering, namely 365/389 (94%). The adjacency of V and P in (a) and (b) guarantees the smallest possible string of words for the recognition and construction of VP and its two constituents (V and PP), see the underlinings in (12):

(12) a. vp[**travels** pp[**to** the city]] b. [[the city **to**]pp **travels**]vp
--- ---
c. vp[**travels** [the city **to**]pp] d. [pp[**to** the city] **travels**]vp
--- ---

The (a) and (b) orders have the highest IC-to-word ratios for the recognition and construction of VP and its two immediate constituents (ICs) V and PP, namely 2/2 = 100%. Non-adjacent V and P in (c) and (d) require longer and less efficient strings for the parsing of phrase structure, namely 2/4 = 50% IC-to-word ratios. That is to say, positioning heads next to one another provides a minimal Phrasal Combination Domain (see (3) above) for the construction of VP and its daughters. Similarly, in the performance data of (5) and (9) orderings are preferred in which heads are adjacent or as close to one another as possible.

Consistent head ordering in grammars can be argued to derive from Minimize Domains, therefore. Conventions of ordering have emerged out of performance preferences, and one and the same principle explains both the preferred conventions of grammar and the preferred structural selections in performance (in languages and structures in which speakers have a choice). MiD can also explain why there are two productive mirror-image types here, head-initial and head-final languages exemplified by (12a) and (b): they are equally good strategies for phrase structure comprehension and production (Hawkins, 2004, pp. 123-126).

According to Dryer (1992) there are systematic exceptions to the preferred consistent head ordering when the category that modifies a head is a single-word item, e.g. an adjective modifying a noun (<u>yellow book</u>). Many otherwise head-initial languages have non-initial heads with the adjective preceding the noun here (e.g. English), many otherwise head-final languages have the noun before the adjective (e.g. Basque). But when the non-head is a branching phrasal category (e.g. an adjective phrase as in English <u>books yellow with age</u>) there are good correlations with the predominant head ordering. Why?

When heads are separated by a non-branching single word, then the difference between the heads V (projecting to VP) and N (projecting to NP) in, say, (13a) and (13b), is short, only one word:

(13) a. vp[**V** [Adj **N**]np]
[read [*yellow book*]]
b. vp[**V** np[**N** Adj]]
[*read* [*book yellow*]]

Hence the MiD preference for noun initiality (and correspondingly for noun-finality in postpositional languages) is significantly less than it is for intervening longer branching phrases, and either less head ordering consistency or no consistency is predicted. That is why English grammar allows (13a), but no NP structure corresponding to [*book* [*yellow with age*]] in which the adjective phrase precedes *book*.

Romance languages have both prenominal and postnominal adjectives, e.g. French *grand homme/homme grand*, but always postnominal adjective phrases just as in English. Similarly, when there is just a one-word difference between competing domains in performance, e.g. in the corpus data of English and Japanese above, both ordering options are generally productive, and so they are too in grammars, as we have just seen.

The Greenbergian word order universals exemplified in (11) above and summarized in Greenberg (1966) and Dryer (1992) support Minimize Domains, therefore (see Hawkins, 2004, 2014). Minimize Domains also predicts departures from consistent ordering and head adjacency in certain cases, as we have seen. There are also certain conflicts between MiD and other efficiency processing principles, e.g. Fillers before Gaps, which result in NRel in certain (non-rigid) OV languages, see the Maximize On-line Processing principle of Hawkins (op cit) and section 6 below. All of these data provide empirical support for a functional explanation for linear ordering rules in the grammars of the world's languages, in terms of ease and efficiency of on-line processing.

They also support the following general hypothesis about the relationship between performance and grammars:

(14) Performance-Grammar Correspondence Hypothesis (PGCH) [Hawkins, 2004, p. 3]
Languages have conventionalized grammatical properties 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.

In other words, languages have conventionalized or 'fixed' in their grammars the same kinds of preferences and principles that we see in performance, especially in performance data from languages in which speakers have alternatives to choose from.

5. Frequency Effects and Minimize Forms

The basic premise of Minimize Forms is that the processing of linguistic forms (phonemes, morphemes, words and phrases) and their conventionalized property assignments requires effort. Minimizing the forms needed for property assignments is efficient since it reduces that effort by fine-tuning it to information that is already active in processing through accessibility, high frequency, and inferencing strategies of various kinds. It is defined in (15):

(15) Minimize Forms (MiF) [Hawkins, 2004, p. 38]
The human processor prefers to minimize the formal complexity of each linguistic form

F (its phoneme, morpheme, word or phrasal units) 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.

MiF is visible in two sets of variation data across and within languages. The first involves complexity differences between surface forms (morphology and syntax), with preferences for minimal expression (e.g. zero morphemes) in proportion to their frequency of occurrence and hence ease of processing through degree of expectedness (cf. Levy, 2008; Jaeger, 2006). For example, singular number for nouns is much more frequent than plural, absolutive case is more frequent than ergative. Correspondingly singularity on nouns is expressed by shorter or equal morphemes, often zero, almost never by more, as in English cat-0 vs. cat-s. Similarly absolutive case is frequently zero, while ergative is explicitly marked (Comrie, 1978).

A second data pattern captured in MiF involves the number and nature of grammatical distinctions that languages conventionalize. The preferences are again in proportion to their efficiency, including frequency of use, and the result is numerous 'hierarchies' such as Greenberg's (1966) morphological or feature hierarchies. Grammatical hierarchy rankings such as Singular > Plural > Dual > Trial/Paucal (see (18) below) correspond to a frequency/ease of processing ranking, with higher positions receiving less or equal formal marking and more or equal unique forms for the expression of that category alone.

The principle of Minimize Forms leads to some testable predictions that are exemplified in detail in Hawkins (2004), and defined in (16) and (17):

(16) Form Minimization Prediction 1
The formal complexity of each form F is reduced in proportion to the frequency of that F and/or the processing ease of assigning a given property P to a reduced F (e.g. to zero).

(17) Form Minimization Prediction 2
The number of unique F:P pairings in a language is reduced by grammaticalizing or lexicalizing a given F:P in proportion to the frequency and preferred expressiveness of that P in performance.

Some of Greenberg's (1966) hierarchies are shown in (18) (with some reformulations and revisions in the publications given in square brackets):

(18) Singular > Plural > Dual > Trial/Paucal (for number) [Greenberg, 1966; Croft, 2003]
Nominative/Absolutive > Accusative/Ergative > Dative > Other (for case marking) [Primus, 1999]
Masculine, Feminine > Neutral (for gender) [Hawkins, 2004]
Positive > Comparative > Superlative [Greenberg, 1966]

Greenberg (1966) pointed out that these kinds of morphological hierarchies define performance frequency rankings for the relevant properties in each domain. The frequencies of number inflections on nouns in a corpus of Sanskrit, for example, were those given in (19):

(19) Singular = 70.3%; Plural = 25.1%; Dual = 4.6%

By MiF Prediction 1 (in (16)) we therefore expect that for each hierarchy H the amount of formal marking (i.e. phonological and morphological complexity) will be greater or equal down each hierarchy position. The kinds of observations that led Greenberg to this insight about the asymmetrical distribution of forms down these hierarchies can be exemplified with data from Manam, an Austronesian language described by Lichtenberk (1983):

(20) 3rd Singular suffix on nouns = 0 (Manam)
3rd Plural suffix = -di,
3rd Dual suffix = -di-a-ru
3rd Paucal = -di-a-to

The amount of formal marking increases from singular to plural, and from plural to dual, and is equal from dual to paucal, all in accordance with the hierarchy prediction.

By MiF Prediction 2 we see that grammars prioritize categories for unique formal expression in proportion to their relative frequency and preferred expressiveness, e.g. Singular and Plural before Dual before Trial and Paucal (see (18)). The frequencies of these categories in grammars match their relative frequencies in performance, therefore.

Hence, the surface forms of the morphology are reduced, in proportion to frequency and/or ease of processing (by Form Minimization Prediction 1 in (16)). And lexical and grammatical categories are given priority for unique formal expression, in proportion to frequency and/or preferred expression (by Form Minimization Prediction 2 in (17)). The result of both is more minimal forms in proportion to frequency/ease of processing/ preferred expressiveness, i.e. fewer and shorter forms for the expression of the speakers' preferred meanings in performance.

6. Conclusions

What all of these examples of performance-grammar correspondences suggest (see (14) above) is that language processing can help us understand grammars better. This particular function of language, the clear, easy and efficient production and comprehension of language in real time, has shaped the grammatical conventions of languages, i.e. their forms and formal properties.

We can now give an explanation for what has been simply observed and stipulated

so far in grammatical models, e.g. the existence of a head ordering parameter, with head-initial (VO) and head-final (OV) languages being roughly equally productive: the reason we propose is that they are equally efficient for processing whether adjacent heads occur on the left of their sisters (as in English), or on the right (Japanese), on account of Minimize Domains (see (1) above).

This approach enables us to understand exceptions to proposed universals, involving e.g. different orderings for single-word versus phrasal modifiers of heads, as well as grammatical patterns such as center-embedding hierarchies that no grammatical model can explain (see Hawkins, 2004, pp. 128-31). It can also explain a host of grammatical patterns involving the formal marking of categories and their asymmetrical distribution, as predicted by Minimize Forms (see (15) above). A third general efficiency principle, in addition to Minimize Domains and Minimize Forms, that is illustrated and defended at length in Hawkins (2004) and (2014) is Maximize On-line Processing. It is summarized here without further comment in (21):

(21) 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) 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.

In other words, grammarians can benefit from the inclusion of processing ideas in their theories and descriptions of conventionalized grammatical data. The approach illustrated in this paper and in my 1994, 2004, and 2014 books is empirically driven by clear surface patterns from language performance and from cross-linguistic grammars expressed in largely theory-general and theory-neutral terms. It provides a way forward for those of us who are interested in answering the question posed at the beginning of this paper: "has function shaped the forms of language?" In this context "has clear, easy, efficient processing left its mark in grammars?".

The answer, I submit, is a resounding "yes". The processing function has profoundly influenced the forms and formal rules of grammars. The empirical method exemplified here enables us to address and gain clarity on the more linguistic and grammatical question about the relationship between form and function without necessarily first solving all the big theoretical issues about processing that are currently being debated in psycholinguistics.

References

Bybee, J. L. (2007). *Frequency of use and the organization of language*. Oxford: Oxford University Press.

Bybee, J. L. (2010). *Language, usage and cognition*. Cambridge: Cambridge University Press.

Choi, H. W. (2007). Length and order: A corpus study of Korean dative-accusative construction. *Discourse and Cognition, 14*, 207-227.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Comrie, B. (1978). Ergativity. In W. P. Lehmann (Ed.), *Syntactic typology: Studies in the phenomenology of language* (pp. 329-394). Austin: University of Texas Press.

Croft, W. (2003). *Typology and universals* (2nd ed.). Cambridge: Cambridge University Press.

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

Filipovic, L., & Hawkins, J. A. (2013). Multiple factors in second language acquisition: The CASP model. *Linguistics, 51*(1), 145-176.

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

Gibson, E. (2000). The Dependency Locality Theory: A distance-based theory of linguistic complexity. In Y. Miyashita, P. Marantz, & W. O'Neil (Eds.), *Image, language, brain* (pp. 95-112). Cambridge, MA: MIT Press.

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

Greenberg, J. H. (1966). *Language universals with special reference to feature hierarchies*. The Hague: Mouton.

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

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

Hawkins, J. A. (2000). The relative ordering of prepositional phrases in English: Going beyond manner-place-time. *Language Variation and Change, 11*, 231-266.

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

Hawkins, J. A. (2009). Language universals and the performance-grammar correspondence hypothesis. In M. H. Christiansen, C. Collins, & S. Edelman (Eds.), *Language universals* (pp. 54-78). Oxford: Oxford University Press.

Hawkins, J. A. (2014). *Cross-linguistic variation and efficiency*. Oxford: Oxford University Press.

Jaeger, T. F. (2006). *Redundancy and syntactic reduction in spontaneous speech* (Unpublished doctoral dissertation). Stanford University, Stanford, CA.

Jaeger, T. F., & Norcliffe, E. (2009). The cross-linguistic study of sentence production. *Language and Linguistics Compass, 3*(4), 866-887.

Jaeger, T. F., & Tily, H. (2011). On language "utility": Processing complexity and communicative efficiency. *WIREs: Cognitive Science, 2*(3), 323-35.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*, 122-149.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition, 106*, 1126-1177.

Lewis, R., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science, 29*, 375-419.

Lichtenberk, F. (1983). *A grammar of Manam*. Honolulu: University of Hawaii Press.

MacWhinney, B., Malchukov, A., & Moravcsik, E. (Eds.). (2014). *Competing motivations in grammar and usage*. Oxford: Oxford University Press.

Newmeyer, F. J. (1998). *Language form and language function*. Cambridge, MA: MIT Press.

Newmeyer, F. J., & Preston, L. B. (Eds.). (2014). *Measuring grammatical complexity*. Oxford: Oxford University Press.

O'Grady, W. (2005). *Syntactic carpentry: An emergentist approach to syntax*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Primus, B. (1999). *Cases and thematic roles: Ergative, accusative and active*. Tuebingen: Max Niemeyer Verlag.

Stallings, L. M. (1998). *Evaluating heaviness: Relative weight in the spoken production of heavy-NP shift* (Doctoral dissertation). University of Southern California.

Traxler, M. J. (2011). *Introduction to psycholinguistics: Understanding language science*. Oxford: Wiley-Blackwell.

Vassisth, S. (2011). Integration and prediction in head-final structures. In H. Yamashita, et al. (Eds.), *Processing and producing head-final structures* (pp. 349-368). New York: Springer.

Wasow, T. (2002). *Postverbal behavior*. Stanford: CSLI Publications, Stanford University.

Yamashita, H., & Chang, F. (2001). "Long before short" preference in the production of a head-final language. *Cognition, 81*, B45-B55.

Yamashita, H., & Chang, F. (2006). Sentence production in Japanese. In M. Nakayama, R. Mazuka, & Y. Shirai (Eds.), *Handbook of East Asian psycholinguistics* (Vol. 2). Cambridge: Cambridge University Press.

Yamashita, H., Hirose, Y., & Packard, J. L. (Eds.). (2011). *Processing and producing head-final structures*. New York: Springer.

About the author

John A. (Jack) Hawkins (jhawkins@ucdavis.edu) is Distinguished Professor of Linguistics at the University of California, Davis and Emeritus Professor of English and Applied Linguistics at Cambridge University. He has held prior permanent positions at the University of Southern California (Los Angeles), the Max-Planck-Institute for Psycholinguistics (Nijmegen), and the University of Essex (Colchester), and visiting appointments at institutions including UC Berkeley, UCLA, the Free University of Berlin, the University of Potsdam, and the Max-Planck-Institute for Evolutionary Anthropology (Leipzig). He has broad and interdisciplinary interests in the language sciences. His current research programs focus on language universals from a processing perspective, on second language acquisition using electronic learner corpora, and on English and the Germanic language family. He has over 100 publications in these areas, including 12 books authored or edited.