Title

Language Universals and the Performance-Grammar Correspondence Hypothesis

ISBN
9780195305432

Author
Hawkins, John A

Publication Date
2009-03-26

DOI
10.1093/acprof:oso/9780195305432.003.0004

Peer reviewed

4

LANGUAGE UNIVERSALS AND THE PERFORMANCE-GRAMMAR CORRESPONDENCE HYPOTHESIS

John A. Hawkins

4.1. Introduction[1]

In this chapter I explore the kinds of variation-defining universals that Greenberg introduced in his seminal 1963 paper on word order. They took the form of implicational statements: if a language has some property (or set of properties) P, then it also has (or generally has) property Q. For example, if a language has subject-object-verb (SOV) order, as in Japanese, it generally has postpositional phrases ([the movies **to**] **went**), rather than prepositional phrases as in English (**went** [**to** the movies]). I will have much less to say about the other major type of universal, the absolute kind of the form "all languages (or no languages) have property P." These have been at the core of Universal Grammar (UG) within generative theories, and are subdivided into "substantive," "functional," and "formal universals" in Steedman (Chapter 9), who follows Chomsky (1995). When implicational universals were incorporated into generative grammar, in the Government-Binding theory of the 1980s (Chomsky, 1981), they became known as "parameters," and the innateness claimed for the absolute universals (Chomsky, 1965; Hoekstra & Kooij, 1988) was extended to the parameters (Fodor, 2001; Lightfoot, 1991). It was proposed that the child's linguistic environment "triggered" one innate parameter rather than another, based on the data of experience. (See Bever, Chapter 6, for a more recent formulation.)

The distinction between variation-defining and absolute universals has taken center stage recently with the publication of Newmeyer's (2005) book *Possible and Probable Languages*: *A Generative Perspective on Linguistic Typology*. Newmeyer argues (contrary to the position taken by Boeckx and Hornstein, in Chapter 5) that the major parameters proposed, the head ordering parameter, the pro-drop parameter, and so on, have systematic exceptions across languages, are probabilistic, and are not part of UG, which is concerned with defining possible versus impossible languages. Haspelmath (2006) gives a similar critique of parameters. In effect, these authors recognize what Greenberg (1963) first recognized: the majority of his implicational statements hold only "with more than chance frequency," and most of those he formulated as exceptionless have subsequently turned out to have exceptions (Dryer, 1992). Clearly, if these parameters are not correct descriptively, they are not innate either, and the kind of environmental trigger theory for language acquisition built around them fails, if the basic premise fails (the existence of innate parameters).

The question then arises: Where do we go from here in order to better understand crosslinguistic variation? A number of generative theorists are trying to improve the empirical adequacy of earlier predictions. Cinque (2005) is a laudable example, which combines Kayne's (1994) antisymmetry principle with painstaking typological work (but see Steedman [2006] for a critique and an alternative). Another research program, more in line with Newmeyer's (op cit) proposals, is the one I shall illustrate in this chapter. Together with many collaborators, I have been pursuing an empirical and interdisciplinary approach to language universals, comparing variation patterns *within* and *across* languages. That is, we have been examining variation both in usage (performance) and in grammars. This program makes extensive use of generative principles and of typologists' generalizations (Comrie, 1989; Croft, 2003), and integrates them with psycholinguistic models and findings.

There are two reasons why this has proved fruitful. First, a general correlation is emerging: the patterns of preference that one finds in performance in languages possessing several structures of a given type (different word orders, relative clauses, etc.) look increasingly like the patterns found in the fixed conventions of grammars in languages with fewer structures of the same type. Numerous examples will be given in what follows.

Second, if this correlation is even partly correct, it has far-reaching consequences for language universals and for the theory of grammar. It enables us to make predictions from performance data for grammatical conventions, and the grammatical patterns predicted are often unexpected from grammatical considerations alone. It helps us to understand not only why there are patterns across languages, but also why there are exceptions to these patterns and when they occur.

Greenberg (1966) was the first to draw attention to such correlating patterns in his discussion of markedness hierarchies like Singular > Plural > Dual > Trial/Paucal. Morphological inventories across grammars and declining allomorphy provided evidence for these universal hierarchies, while declining frequencies of use in languages with rich inventories suggested not only a correlation with performance but also a possibly causal role for it in the evolution of the grammatical regularities themselves (Greenberg, 1995, pp. 163–164). Givón (1979, pp. 26–31) meanwhile observed that performance preferences in one language, for definite subjects, for example, may correspond to an actual categorical requirement in another. In Hawkins (1990, 1994), I argued that the preferred word orders in languages with choices are those that are productively conventionalized as fixed orders in languages with less freedom. And in my 2004 book I examine many more grammatical areas in a systematic test of the following hypothesis:

(1) *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.

There is a growing awareness of this basic correspondence in many branches of the language sciences. Haspelmath (1999) proposed a theory of diachrony in which usage preferences lead to changing grammatical conventions over time. Bybee and Hopper (2001) document the clear role of frequency in the emergence of grammatical structure. There have been intriguing computer simulations of language evolution, exemplified by Kirby (1999), in which processing preferences of the kind assumed for word order in Hawkins (1990, 1994) are incorporated in the simulation and lead to the emergence of the observed grammatical types after numerous iterations (corresponding to successive generations of language users). There have been developments in Optimality Theory, exemplified by Haspelmath (1999) and Aissen (1999), in which functional motivations of an ultimately processing nature are provided for many of the basic constraints. Stochastic Optimality Theory (Bresnan, Dingare, & Manning, 2001; Manning, 2003) incorporates the preferences of performance ("soft constraints") as well as grammatical conventions ("hard constraints"). Newmeyer (2005) advocates replacing generative parameters with principles derived from language processing, while Phillips (1996) and Kempson, Meyer-Viol, and Gabbay (2001) incorporate the online processing of language into the rules and representations of the grammar.

But despite this growing interest in performance-grammar correspondences, the precise extent to which grammars have been shaped by performance is a matter of intense debate. There are different opinions in the publications cited so far and in the chapters of this volume. In the present context, I shall accordingly focus on the empirical evidence for the PGCH in order to try and convince the next generation of researchers that there is a real generalization here and that it does need to be incorporated into theories of grammatical universals. In the next section (section 4.2), I briefly summarize a range of observed performance-grammar correspondences that support it. I then exemplify the testing of the PGCH in the area of word order (section 4.3), followed by a short discussion of relative clauses (section 4.4). Conclusions and further issues are summarized in section 4.5.

4.2. Examples of Proposed Performance-Grammar Correspondences

The Keenan and Comrie (1977) Accessibility Hierarchy (SU>DO>IO/OBL>GEN; cf. Comrie, 1989) has been much discussed in this context. Grammatical cut-off points in relativization across languages follow the hierarchy, and Keenan and Comrie argued for an explanation in terms of declining ease of processing down the lower positions of the hierarchy. As evidence, they pointed to usage data from languages with many relativizable positions, especially English. The hierarchy correlated both with declining corpus frequencies down the hierarchy and with evidence of increasing processing load and working memory demands under experimental conditions (Diessel & Tomasello, 2006; Hawkins, 1999; Keenan, 1975; Keenan & Hawkins, 1987; cf. section 4.4.1).

More generally, filler-gap dependency hierarchies for relativization and Wh-movement structures across grammars point to increasing complexity in the permitted gap environments. For example, grammatical cut-off points in increasingly complex clause-embedding positions for gaps correspond to declining processing ease in languages with numerous gap-containing environments (including subjacency-violating languages like Akan; Saah & Goodluck [1995]); cf. Hawkins (2004; Chapter 7) and section 4.4.2.

Reverse hierarchies across languages for gaps in simpler relativization domains and resumptive pronouns in more complex environments (Hawkins, 1999) match the performance distribution of gaps to pronouns within languages such as Hebrew and Cantonese in which both are grammatical (in some syntactic positions), gaps being preferred in the simpler, and pronouns in the more complex relatives (Ariel, 1999; Hawkins, 2004; Matthews & Yip, 2003); cf. section 4.4.1.

Parallel function effects (whereby the head of the relative matches the position relativized on) have been shown to facilitate relative clause processing and acquisition (Clancy, Lee, & Zoh, 1986; MacWhinney, 1982; Sheldon, 1974). They also extend relativization possibilities beyond normal constraints holding in languages such as Basque and Hebrew (Aldai, 2003; Cole, 1976; Hawkins, 2004, p. 190).

Declining acceptability of increasingly complex center embeddings, in languages in which these are grammatical, is matched by hierarchies of permitted center embeddings across grammars, with cut-offs down these hierarchies (Hawkins, 1994, pp. 315–321).

(Nominative) subject (S) before (accusative) object (O) ordering is massively preferred in the performance of languages in which both SO and OS are grammatical (Japanese, Korean, Finnish, German) and is also massively preferred as a basic order or as the only order across grammars (Gibson, 1998; Hawkins, 1994; Miyamoto, 2006; Primus, 1999; Tomlin, 1986).

Markedness hierarchies of case (Nom>Acc>Dat>Other), number (Sing>Plur>Dual>Trial), etc., correspond to performance frequency hierarchies in languages with rich morphological inventories (Croft, 2003; Greenberg, 1966; Hawkins, 2004, pp. 64–68).

Performance preferences for subjects that obey the Person Hierarchy (first > second > third) in English (whereby *the boy hit me* is preferably passivized to *I was hit by the boy*) have been conventionalized into a grammatical/ungrammatical distinction in languages such as Lummi (Bresnan, Dingare, & Manning, 2001). Sentences corresponding to *the boy hit me* are ungrammatical in Lummi.

The distinction between zero agreement in local NP environments versus explicit agreement nonlocally in the grammar of Warlpiri matches the environments in which zero and explicit forms are preferred in performance in languages with choices (Hawkins, 2004, p. 160).

I believe these are the tip of a large iceberg of performance-motivated crosslinguistic patterns. And if these correspondences are valid, then the classic picture of the performance-grammar relationship presented in Chomsky (1965) needs to be revised. For Chomsky, the competence grammar was an integral part of a performance model, but it was not shaped by performance in any way:

> Acceptability . . . belongs to the study of performance, . . . The unacceptable grammatical sentences often cannot be used, for reasons having to do . . . with memory limitations, intonational and stylistic factors, . . . and so on. . . . it would be quite impossible to characterize unacceptable sentences in grammatical terms . . . we cannot formulate particular rules of the grammar in such a way as to exclude them.
>
> (Chomsky, 1965, pp. 11–12)

Chomsky claimed (and still claims) that grammar was autonomous and UG was innate (see Newmeyer [1998] for a full summary and discussion of these points). The PGCH in (1) is built on the opposite assumption that grammatical rules *have* incorporated properties that reflect memory limitations and other forms of complexity and efficiency that we observe in performance. This alternative is supported by the correspondences above, and it makes predictions for occurring and nonoccurring grammars, and for frequent and less frequent ones. It accounts for many crosslinguistic patterns that are not predicted by grammar-only theories and for exceptions to those that are predicted. In the next section, I illustrate the PGCH and this research method in greater detail.

4.3. Head Ordering and Adjacency in Syntax

I begin by examining some variation data from English and Japanese in which users have a choice between the adjacency or nonadjacency of certain categories to their heads. It turns out that there are systematic preferences in performance, mirror image ones interestingly between these languages, and an efficiency principle of Minimize Domains is proposed that describes these preferences. I then show that this same principle can be found in the fixed conventions of grammars in languages with fewer options. Specifically, this principle can give us an explanation, derived from language use and processing, for general patterns in grammars, for puzzling exceptions to these patterns, and for grammatically unpredicted data sets involving, for example, hierarchies.

The principle of Minimize Domains is defined at the outset (cf. Hawkins, 2004, p. 31):

(2) *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.

Combination: Two categories, A and B, are in a relation of combination iff they occur within the same syntactic mother phrase or maximal projection (phrasal combination), or if they occur within the same lexical co-occurrence frame (lexical combination).

Dependency: Two categories, A and B, are in a relation of dependency iff the parsing of B requires access to A for the assignment of syntactic or semantic properties to B with respect to which B is zero-specified, or ambiguously or polysemously specified.

4.3.1. *Syntactic MiD Effects in the Performance of Head-Initial Languages*

Words and phrases have to be assembled in comprehension and production into the kinds of groupings that are represented by tree structure diagrams. Recognizing how words and phrases combine together 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 a mother phrase M and its immediate constituent daughters (ICs), making phrasal combination faster. Compare (3a) and (3b):

(3) 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, PP1, and PP2, can be recognized on the basis of five words in (3a) compared with nine in (3b), assuming that (head) categories such as P immediately project to mother nodes such as PP, enabling the parser to construct and recognize them online. For comparable benefits within a production model, cf. Hawkins (2004, p. 106).[2]

Minimize Domains predicts that Phrasal Combination Domains (PCDs) should be as short as possible, and that the degree of this preference should be proportional to the minimization difference between competing orderings. This principle (a particular instance of Minimize Domains) is called Early Immediate Constituents (EIC):

(4) *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 nonterminals over the terminals) on the basis of which the processor can construct M and its ICs.

(5) *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—for example, short before long PPs in (3). These orders will have higher "IC-to-word ratios," that is, 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 (3a) is 3/5 or 60% (5 words required for the recognition of 3 ICs). The comparable ratio for (3b) is 3/9 or 33% (9 words required for the same 3 ICs).

Structures like (3) 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 (PPs = shorter PP, PPL = longer PP):

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

Numerous other structures reveal the same weight preference in English (e.g., Heavy NP Shift); cf. Hawkins (1994, p. 183), Wasow (1997, 2002), and Stallings (1998).

A possible explanation for the distribution in (6) can be given in terms of reduced simultaneous processing demands in working memory. If, in (3a), 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, and there will be fewer additional (phonological, morphological, syntactic, and semantic) decisions that need to be made simultaneously with this one, and less demands on working memory; (3a) is therefore more efficient. More generally, we can hypothesize that the processing of all syntactic and semantic relations prefers minimal domains, which is what MiD predicts (Hawkins, 2004).[3]

4.3.2. Minimal Domains for Lexical Combinations and Dependencies

A PCD is a domain for the processing of a syntactic relation of phrasal combination or sisterhood. Some of these sisters contract additional relations of a semantic and/or syntactic nature, of the kind grammatical models try to capture in terms of verb-complement (rather than verb-adjunct) relations, such as *count on your father* versus *play on the playground* (place adjunct). Complements are listed in the lexical entry for each head, and the processing of verb–complement relations should also prefer minimal domains, by MiD, (2).

(7) *Lexical Domain* (LD)
The LD for assignment of a lexically listed property P to a lexical item L consists of the smallest possible string of terminal elements (plus their associated syntactic and semantic properties) on the basis of which the processor can assign P to L.

One practical problem here is that the complement/adjunct distinction is a multi-factor one covering different types of combinatorial and dependency relations, obligatoriness versus optionality, etc., and is not always clear (cf. Schütze & Gibson, 1999). Hawkins (2000, 2001) proposes the following entailment tests as a way of defining PPs that are lexically listed:

(8) Verb Entailment Test: Does [V, {PP1, PP2}] entail V alone or does V have a meaning dependent on either PP1 or PP2? Example: *The man waited for his son in the early morning* entails *the man waited*; *the man counted on his son in his old age* does not entail *the man counted*.

(9) Pro-Verb Entailment Test: Can V be replaced by some general Pro-verb or does one of the PPs require that particular V for its interpretation? Example: *The boy played on the playground* entails *the boy did something on the playground*, but *the boy depended on his father* does not entail *the boy did something on his father*.

If V or P is dependent on the other by these tests, then the PP is lexically listed, that is, dependency is used as a sufficient condition for complementhood and lexical listing. The PPs classified as independent are (mostly) adjuncts or unclear cases.

When there was a dependency between V and just one of the PPs, then 73% (151/206) had the interdependent PP (Pd) adjacent to V, that is, their LDs were minimal. Recall that 82% had a short PP adjacent to V preceding a longer one in (2), that is, their PCDs were minimal. For PPs that were *both* shorter *and* lexically dependent, the adjacency rate to V was 96%, which was (statistically) significantly higher than for each factor alone.

We can conclude that the more syntactic and semantic relations whose domains are minimized in a given order, the greater is the preference for that order: multiple preferences result in a stronger adjacency effect when they reinforce each other, as predicted by MiD (2). MiD also predicts a stronger adjacency preference within each processing domain in proportion to the minimization difference between competing sequences. For PCDs, this difference is a function of the relative weights of the sisters; cf. (6). For LDs, it is a function of the absolute size of any independent PP (Pi) that could intervene between the verb and the interdependent PP (Pd) by the entailment tests, thereby delaying the processing of lexical co-occurrences.

(10) n = 206	Pi = 2–3 words	:4–5	:6–7	:8+
[V Pd Pi]	59% (54)	71% (39)	93% (26)	100% (32)
[V Pi Pd]	41% (37)	29% (16)	7% (2)	0% (0)

Multiple preferences have an additive adjacency effect when they work together, but they result in exceptions to each when they pull in different directions. Most of the 58 long-before-short sequences in (6) involve some form of lexical dependency between V and the longer PP (Hawkins, 2000). Conversely, V and Pd can be pulled apart by EIC in proportion to the weight difference between Pd and Pi (Hawkins, 2004, p. 116).

4.3.3. *MiD Effects in Head-Final Languages*

Long before short orders provide minimal PCDs in head-final languages in which constructing categories (V, P, Comp, case particles, etc.) are on the right. For example, if the direct object in Japanese is a complement clause headed by the complementizer *to*, as in (11), the distance between the complementizer and other constituents of the matrix clause, the subject *Mary ga* and the verb *it-ta*, will be very short in (11b), just as short as it is in the mirror image English translation *Mary said that* . . . . Hence the Phrasal Combination Domain for the matrix clause in (11b) is minimal. In (11a), by contrast, with the center-embedded complement clause, this PCD proceeds all the way from *Mary ga* to *it-ta*, and is much longer.

(11) a. Mary ga [[kinoo John ga kekkonsi-ta to]s it-ta]vp
Mary NOM yesterday John NOM married that said
Mary said that John got married yesterday.
b. [kinoo John ga kekkonsi-ta to]s Mary ga [it-ta]vp

A preference for (11b) is accordingly predicted in proportion to the relative weight difference between subject and object phrases. By similar reasoning, a long-before-short preference is predicted for [{NPo, PPm} V] structures in Japanese, in alternations such as (12) (with *-o* standing for the accusative case particle, and PPm for a postpositional phrase with a head-final postposition):

(12) a. (Tanaka ga) [[Hanako kara]pp [sono hon o]np katta]vp
Tanaka NOM Hanako from that book ACC bought,
"Tanako bought that book from Hanako"
b. (Tanaka ga) [[sono hon o]np [Hanako kara]pp katta]vp

Relevant corpus data were collected by Kaoru Horie and are reported in Hawkins (1994, p. 152). Letting ICs and ICL stand for shorter and longer ICs, respectively (i.e., with weight as the crucial distinction rather than phrasal type), these data are summarized in (13) (excluding the phrases with equal weights):

(13) n = 153	ICL>ICs by 1–2 words	by 3–4	by 5–8	by 9+
[ICs ICL V]	34% (30)	28% (8)	17% (4)	9% (1)
[ICL ICs V]	66% (59)	72% (21)	83% (20)	91% (10)

These data are the mirror image of those in (6): the longer IC is increasingly preferred to the left in the Japanese clause, whereas it is increasingly preferred to the right in English. This pattern has since been corroborated in experimental and further corpus data by Yamashita and Chang (2001), and it underscores an important principle for psycholinguistic models. The directionality of weight effects depends on the language type. Heavy phrases shift to the right in English-type (head-initial) structures, and to the left in Japanese-type (head-final) structures; cf. Hawkins (1994, 2004) for extensive illustration and discussion.

In (14), the data of (13) are presented for both phrasal type (NPo versus PP) and relative weight:

(14)

	NPo>PPm by			NPo=PPm	PPm>NPo by		
n = 244	5+	3–4	1–2		1–2	3–8	9+
[PPm NPo V]	21% (3)	50% (5)	62% (18)	66% (60)	80% (48)	84% (26)	100% (9)
[NPo PPm V]	79% (11)	50% (5)	38% (11)	34% (31)	20% (12)	16% (5)	0% (0)

Notice the preferred adjacency of a direct object NPo complement to V when weight differences are equal or small in (14). This interacting preference is plausibly a consequence of the fact that NPs are generally complements and in a lexical combination with V, whereas PPs are either adjuncts or complements, mostly the former; cf. section 4.3.2.

4.3.4. Greenberg's Word Order Correlations and Other Domain Minimizations

Grammatical conventions across languages reveal the same degrees of preference for minimal domains. The relative quantities of languages reflect the preferences, as do hierarchies of co-occurring word orders. An efficiency approach can also explain exceptions to the majority patterns and to grammatical principles such as consistent head ordering.

Let us return to the implicational universal with which I began this chapter (section 4.1). Greenberg (1963) examined alternative verb positions across languages and their correlations with prepositions and postpositions in phrases corresponding to (15):

(15) a. vp[went pp[to the movies]] b. [[the movies to]pp went]vp
        --------   --------
     c. vp[went [the movies to]pp] d. [pp[to the movies] went]vp
        ----------------   ----------------

(15a) is the English order, (15b) is the Japanese order, and these two sequences, with adjacent lexical heads (V and P), are massively preferred in language samples over the inconsistently ordered heads in (15c) and (15d). (16) summarizes the distribution in the database of Dryer's (1992) paper on the "Greenbergian correlations" (Hawkins, 2004, p. 124):

(16) 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%)
     Preferred (16a) + (b) = 365/389 (94%)

The adjacency of V and P guarantees the smallest possible string of words, indicated by the underlinings in (15), for the recognition and construction of VP and of its two immediate constituents (ICs), namely, V and PP. Non-adjacent V and P in (15c) and (15d) require longer and less efficient strings for the parsing of phrase structure. That is, adjacency provides a minimal Phrasal Combination Domain for the construction of VP and its daughters, of the same kind we saw in the performance preferences of sections 4.3.1–4.3.3.

Consistent head ordering in grammars can be argued to derive from Minimize Domains (2), therefore. Conventions of ordering have emerged out of performance preferences, and one and the same principle can explain 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 (15a) and (b), respectively: they are equally good strategies for phrase structure comprehension and production (Hawkins, 2004, pp. 123–126).

Purely grammatical approaches can also define a head ordering parameter (cf. Newmeyer, 2005, p. 43 and Haspelmath, 2006, for full references), and Svenonius (2000, p. 7) states that this parameter is "arguably not a function of processing." It is certainly possible that this is an autonomous principle of grammar with no basis in performance. But how do we argue for or against this?

A classic method of reasoning in generative grammar has always involved capturing significant generalizations and deriving the greatest number of observations from the smallest number of principles. An autonomous head ordering principle would fail to capture the generalization that both grammatical and performance data fall under Minimize Domains. The probabilistic and preferential nature of this generalization is also common to both. Moreover, many other ordering universals point to the same preference for small and efficient Phrasal Combination Domains, for example, in noun-phrase-internal orderings corresponding to (17) in English:

(17) np[bright students s'[that Mary will teach]]

(17') np[Adj N s'[C S]] C = the category that constructs S': e.g., relative pronoun, complementizer, subordinating affix or particle, participial marking on V, etc. (Hawkins, 1994, pp. 387–393)

There are 12 logically possible orderings of Adj, N, and S' (ordered [C S] or [S C]). Just four of these have minimal PCDs for the NP (100% IC-to-word ratios), all of them with adjacent Adj, N, and C, namely, [N Adj [C S]] (Romance), [Adj N [C S]] (Germanic), [[S C] N Adj] (Basque), and [[S C] Adj N] (Tamil). These four account for the vast majority of languages, while a small minority of languages are distributed among the remaining eight in proportion to their IC-to-word ratios measured on-line (Hawkins, 1990, 1994, 2004). There appears to be no straightforward grammatical account for this distribution of occurring versus non-occurring and preferred versus less preferred grammars. The distribution does correlate with degrees of efficient processing in NP Phrasal Combination Domains, however.

4.3.5. Explaining Grammatical Exceptions and Unpredicted Patterns

Further support for the Minimize Domains explanation for head ordering comes from the grammars with exceptional head orderings. Dryer (1992) points out that there are systematic exceptions to Greenberg's correlations when the category that modifies a head is a single-word item, for example, an adjective modifying a noun (*yellow book*). Many otherwise head-initial languages have noninitial heads here (English is a case in point), and many otherwise head-final languages have noun before adjective (e.g., Basque). But when the non-head is a branching phrasal category (e.g., an adjective phrase as in English, *books yellow with age*), there are good correlations with the predominant head ordering. Why should this be?

When a head category like V has a phrasal sister, for example, PP in {V, PP}, then the distance from the higher head to the head of the sister will be very long when heads are inconsistently ordered and are separated by a branching phrase (e.g., vp[V [NP P]pp] in [15c]. An intervening phrasal NP between V and P makes the PCD for the mother VP long and inefficient compared with the consistently ordered counterpart (15a), vp[V pp[P NP]], in which just two words suffice to recognize the two ICs. But when heads are separated by a nonbranching single word, then the difference between, say, vp[V np[N Adj]] and vp[V [Adj N]np] is short, only one word. Hence, the MiD preference for noun initiality (and for noun finality in postpositional languages) is significantly less than it is for intervening branching phrases, and either less head ordering consistency or no consistency is predicted. When there is just a one word difference between competing domains in performance, cf. (6), both ordering options are generally productive, and so too in grammars.

MiD can also explain numerous patterns across grammars that do not follow readily from grammatical principles alone. Hierarchies of permitted center-embedded phrases are a case in point. For example, in the environment pp[P np[__ N]], we have the following center-embedding hierarchy (Hawkins, 1983):

(18)	Prepositional languages:	DemN	49%	NDem	51%
		AdjN	32%	NAdj	68%
		PosspN	12%	NPossp	88%
		RelN	1%	NRel	99%

As the aggregate size and complexity of nominal modifiers increases (relative clauses exceeding possessive phrases, which in turn exceed single-word adjectives), the distance between P and N increases in the prenominal order and the efficiency of the PCD for PP declines compared with postnominal counterparts.[4] As efficiencies decline, the relative frequencies of prenominal orders in conventionalized grammatical rules declines also.

4.3.6. Minimal Domains for Complements and Heads in Grammars

Complements prefer adjacency over adjuncts in the basic orders of numerous phrases in English and other languages and are generated in a position adjacent to their heads in the grammars of Jackendoff (1977) and Pollard and Sag (1987). Tomlin's (1986) verb–object bonding discussion provides cross-linguistic support for this by pointing to languages in which it is impossible or dispreferred for adjuncts to intervene between a verbal head and its DO complement.

Why should complements prefer adjacency in grammars when there are basic ordering conventions? The reason, I suggest, is the same as the one I gave for the preferred orderings of complements (Pd) in performance in section 4.3.2. There are more combinatorial and/or dependency relations linking complements to their heads than linking adjuncts. Complements are listed in a lexical co-occurrence frame defined by, and activated by, a specific head (e.g., a verb); adjuncts are not so listed and occur in a wide variety of phrases with which they are semantically compatible (Pollard & Sag, 1994). The verb is regularly lexically dependent on its DO, not on an adjunct phrase: compare the different senses of "run" in *run the race/the water/the advertisement* (*in the afternoon*); cf. Keenan (1979). A direct object receives a theta-role from V, typically a subtype of Dowty's (1991) Proto-Patient; adjuncts don't get theta-roles. The DO is also syntactically required by a transitive V, whereas adjuncts are not syntactically required sisters. Processing these lexical co-occurrence relations favors minimal Lexical Domains (7).

4.4. Relative Clauses

Relative clauses have been well researched across grammars, and they are now receiving increasing attention in performance; so we can begin to compare the two sets of data in a further test of the PGCH (1). Relative clauses may exhibit a "gap" or a "resumptive pronoun" strategy (in Hebrew structures corresponding to *the students* [*that I teach* (*them*)]), or a structure with or without a relative pronoun (in English, cf. *the students* [(*whom*) *I teach*]). One of these strategies can be "fixed" or "conventionalized" in certain environments, whereas there can be optionality and variation in others. The issue is then whether the fixed conventions of grammars match the preferred variants of performance.

4.4.1. Gaps versus Pronouns

The selection from the variants in performance exhibits patterns: the retention of the relative pronoun in English is correlated (inter alia) with the degree of separation of the relative from its head; cf. Quirk (1957). The Hebrew gap has been shown to be favored with smaller distances between filler and gap (Ariel, 1999): that is, (19a) is significantly preferred over (19b) with a resumptive pronoun, while the pronoun becomes productive when "filler-gap domains" (Hawkins, 1999, 2004) would be larger, as in (20).

(19) a. Shoshana hi ha-isha*i* [she-nili ohevet 0*i*] (Hebrew)
Shoshana is the-woman that-Nili loves
b. Shoshana hi ha-isha*i* [she-nili ohevet ota*i*]
that-Nili loves her

(20) Shoshana hi ha-isha*i* [she-dani siper she-moshe rixel she-nili ohevet ota*i*]
Shoshana is the-woman that-Danny said that-Moshe gossiped that-Nili loves her

The distribution of the fixed variants across grammars also reveals patterns. In simple relative clauses in Cantonese, in which the subcategorizing verb would be adjacent to the head of the relative, a resumptive pronoun is ungrammatical (21b). In the more complex environment of (22) (with an intervening embedded VP), both gaps and resumptives occur (Matthews & Yip, 2003)[5]:

(21) a. [Ng05 ceng2 0*i*] g02 di1 pang4jau5*i* (Cantonese)
I invite those CL friend
"friends that I invite"
b. *[Ng05 ceng2*i* keoi5dei6*i*)] g02 di1 pang4jau5*i*
I invite them those CL friend

(22) [Ng05 ceng2 (keoi5dei6*i*) sik6-faan6] g02 di1 pang4jau5*i*
I invite (them) eat-rice those CL friend
"friends that I invite to have dinner"

More generally, the distribution of gaps to pronouns follows the Keenan and Comrie (1977) Accessibility Hierarchy (SU>DO>IO/OBL>GEN). This increasing preference for pronouns down the hierarchy provides a further piece of evidence for their claim that the AH correlates with declining ease of processing. Hawkins (1999, 2004) argued that there are indeed more complex domains for relative clause processing down the AH, measured in terms of syntactic node size and other correlated measures, and he argues further that resumptive pronouns minimize the lexical domains for argument processing, resulting in more efficient structures overall when relativization environments are complex.

In (19b) and (20), for example, the pronoun provides a local argument, *ota* (her), for lexical processing of *ohevet* (loves), whereas in (19a), lexical processing needs to access the more distant head *ha-isha* (woman) in order to assign a direct object to *loves*. The larger the distance between the subcategorizor and the relative clause head, the less minimal this Lexical Domain becomes, and the more efficient the copy pronoun becomes.

4.4.2. *Relative Clause Hierarchies*

A question that arises from the Accessibility Hierarchy is: What other grammatical hierarchies can be set up for relative clauses on the basis of cross-linguistic data, and do their ranked positions lend themselves to an account in terms of processing complexity? In Hawkins (1999, 2004), I proposed the following clause-embedding hierarchy for gaps:

(23) infinitival (VP) complement>finite (S) complement > S within a complex NP

Relativization cuts off down this hierarchy in selected languages, much the way it does down the AH. Some languages exemplifying the cut-off points are summarized in (24):

(24)	Infinitival (VP) complement:	Swedish, Japanese, English, French, German, Russian;
	Finite (S) complement:	Swedish, Japanese, English, French;
	S within complex NP:	Swedish, Japanese.

Standard German exhibits ungrammaticalities for relative clause gaps in finite complements; cf. (25). Corresponding gaps in infinitival complements such as (26) are grammatical (see Kvam, 1983, and Comrie, 1973, for similar data from Russian):

(25) *Der Tierpark$_i$ [den$_i$ ich vermute s[dass alle Deutschen kennen 0$_i$]] heisst . . . (German)
the zoo which I suppose that all Germans know is-called

(26) das Buch$_i$ [das$_i$ ich vp[0$_i$ zu finden] versucht hatte]/[das$_i$ ich versucht hatte [0$_i$ zu finden]]
the book which I to find tried had / which I tried had to find
"the book which I had tried to find"

English and French cut off at gaps within complex NP clauses, exemplified by (27) from English. This sentence contrasts with its Swedish counterpart, (28), in which the corresponding complex NP structure is completely grammatical (Allwood, 1982):

(27) *I was looking for a bone$_i$ [which$_i$ I saw np[a dog s[that was gnawing on 0$_i$]]]

(28) ett ben$_i$ [som$_i$ jag ser np[en hund s[som gnager på 0$_i$]]] (Swedish)
a bone which I see a dog which is-gnawing on

Corresponding to this last cut-off point, Saah and Goodluck (1995) have presented interesting experimental data from Akan (aka Twi), a language in which there is no grammatical subjacency condition outlawing gaps in complex NPs. Speakers nonetheless showed greater processing difficulty for gaps within a complex NP compared with an embedded finite clause; these processing data matched the grammatical cut-off data on the clause-embedding hierarchy (24).

4.5. Conclusions and Further Issues

The data of this chapter have shown that there are clear parallels between performance variation data and grammatical universals of the variation-defining kind. Hence, any proposed principles that apply to grammars only, such as innate parameters (Chomsky, 1981), are missing a significant generalization. One common principle evident in both is Minimize Domains (2). There is a correlation between the adjacency preferences of performance, in sections 4.3.1–4.3.3, and the adjacency conventions of grammars, sections 4.3.4–4.3.6. Further correlations between performance and grammatical variation were summarized in sections 4.2 and 4.4.

These correlations support the Performance-Grammar Correspondence Hypothesis in (1). The major predictions of the PGCH that are systematically tested in Hawkins (2004) are the following:

(29) *Grammatical predictions of the PGCH* (Hawkins, 2004)

(a) 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.

(b) 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 cutoff points and declining frequencies of languages).

(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.

We have seen in this chapter that such predictions are widely supported. Hence, principles of performance provide an explanation for variation-defining universals. Minimize Domains explains the Greenbergian correlations in (16) and other ordering patterns. It explains why there are two productive language types, head initial and head final: they are both equally efficient according to Minimize Domains (Hawkins, 1994, 2004). It explains puzzling exceptions to consistent head ordering involving single-word versus multi-word modifiers of heads (section 4.3.5). It also explains cross-linguistic patterns that are not predicted by grammatical principles alone, such as hierarchies of increasingly complex center embeddings in (18), and reverse hierarchies for some data versus others. For example, gap relatives cut off down the Keenan-Comrie Accessibility Hierarchy (if a gap occurs low on the hierarchy, it occurs all the way up), whereas resumptive pronouns follow an "if high, then low" pattern (section 4.4.1, and Hawkins, 1999).

Based on this evidence, I conclude, with Newmeyer (2005), that performance and processing must play a central role in any theory of grammatical variation and of variation-defining language universals. The PGCH gives good descriptive coverage. It also provides answers to explanatory questions that are rarely raised in the generative literature, such as the following: Why should there be a head ordering principle defining head-initial and head-final language types (Hawkins, 1990, 1994)? Why are there heads at all in phrase structure (Hawkins, 1993, 1994)? Why are some categories adjacent and others not (Hawkins, 2001, 2004)? Why is there a subjacency constraint, and why is it parameterized the way it is (Hawkins, 1999, 2004)?

They can be asked, and informative answers can be given, in the framework proposed here. The basic empirical issue involves conducting a simple test: Are there, or are there not, parallels between universal patterns across grammars, and patterns of preference and processing ease within languages? The data of this chapter suggest that there are, and the descriptive and explanatory benefits for which I have argued then follow.

Two further common principles of performance and grammar from Hawkins (2004) are summarized here without further comment:

(30) *Minimize Forms* (MiF)
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.

(31) *Maximize Online 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(nline) 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.

Let me end this chapter with some general remarks on further issues. I distinguished in section 4.1 between variation-defining universals and absolute universals, and the data and discussion have been concerned with the former. I have argued that innate grammatical knowledge cannot be the ultimate explanation for them, but notice that it is still plausible to think in terms of Elman et al. (1996) "architectural innateness" as constraining the data of performance, which then evolve into conventions of grammar. The architectural innateness of the human language faculty enters into grammars indirectly in this way. Absolute universals can also be innately grounded as a result of processing constraints on grammars. When complexity and efficiency levels are comparable and tolerable, we get the variation between grammars that we have seen. But within and beyond certain thresholds, I would expect universals of the kind "all languages have X" and "no languages have X," respectively, as a result of processability interacting with the other determinants of grammars. The PGCH is no less relevant to absolute universals, therefore, with the extremes of simplicity/complexity and (in)efficiency being inferrable from actually occurring usage data. Systematic exploration of this idea is required in order see to what extent absolute universals can be explained through processing as well.

There can also be innate grammatical and representational knowledge of quite specific properties of the kind summarized in Pinker and Jackendoff's (Chapter 7) response to Hauser, Chomsky, and Fitch (2002). Much of phonetics, semantics, and cognition is presumably innately grounded, and there are numerous properties unique to human language as a result. See Newmeyer (2005) for the role of conceptual structure in shaping absolute universals, and also Bach and Chao (Chapter 8) for a discussion of semantically based universals.

The precise causes underlying the observed preferences in performance require more attention than I have given them here, and indeed much of psycholinguistics is currently grappling with this issue. To what extent do the preferences result from parsing and comprehension, and to what extent are they production-driven? See, for example, Wasow (2002) and Jaeger and Wasow (2005) for discussion of the different predictions made by production- versus comprehension-based theories for some of the data of this chapter. In addition, what is the role of frequency sensitivity and of prior learning in on-line processing? (e.g., Reali & Christiansen, 2007a, b).

A performance explanation for universals has consequences for learning and for learnability since it reduces the role of an innate grammar. UG is no longer available in the relevant areas (head ordering, subjacency, etc.) to make up for the claimed poverty of the stimulus and to solve the negative evidence problem (Bowerman, 1988). The result is increased learning from positive data, something that Tomasello (2003), connectionist modelers like MacDonald (1999), and also linguists like Culicover (1999) have been arguing for independently. These converging developments enable us to see the data of experience as less impoverished and more learnable than previously thought. The grammaticality facts of Culicover's book, for example, pose learnability problems that are just as severe as those for which Hoekstra and Kooij (1988) invoke an innate UG, yet Culicover's data involve language-particular subtleties of English that cannot possibly be innate (*the student is likely to pass the exam* versus **the student is probable to pass the exam*). See Hawkins (2004, pp. 272–276) for further discussion of these issues.

The explanation for cross-linguistic patterns that I have proposed also requires a theory of diachrony that can translate the preferences of performance into fixed conventions of grammars. Grammars can be seen as complex adaptive systems (Gell-Mann, 1992), with ease of processing driving the adaptation in response to prior changes. But we need to better understand the "adaptive mechanisms" (Kirby, 1999) by which grammatical conventions emerge out of performance variants. How do grammatical categories and the rule types of particular models end up encoding performance preferences? And what constraints and filters are there on this translation from performance to grammar? I outlined some major ways in which grammars respond to processing in Hawkins (1994, pp. 19–24) (by incorporating movement rules applying to some categories rather than others, defining certain orderings rather than others, constraining the applicability of rules in certain environments, etc.), and I would refer the reader to that discussion. How are these different rules then selected by successive generations of learners, and even by the same generation over time? I refer the reader here to Haspelmath (1999) and to Hurford (Chapter 3).

Key Further Readings

A non-technical introduction to the PGCH presented in this chapter is given in the first two chapters of Hawkins (2004). Chapters 5–8 of that book provide more detailed justification for the ideas presented here. An up-to-date introduction oriented to the concerns of typologists can be found in my chapter, "Processing efficiency and complexity in typological patterns," to appear in the *Oxford Handbook of Language Typology*, edited by Jae Jung Song. I give an introduction for psychologists, "Processing Typology and why psychologists need to know about it," in *New Ideas in Psychology* 25: 87–107 (2007). Greenberg's classic (1963) paper referenced below on the order of meaningful elements is essential reading for the study of variation-defining universals, and Newmeyer's (2005) book on possible and probable languages gives a good summary of both generative and typological approaches to these universals and a generative perspective that complements the typological and psycholinguistic perspective presented here. For introductions to typology, see Comrie (1989) and Croft (2003).

Notes

1 The following abbreviations are used in this chapter—Acc: accusative case; Adj: adjective; AH: Accessibility Hierarchy; C: category that constructs S'; CL: classifier; Comp: complementizer; Dat: dative case; Dem: demonstrative determiner; DO: direct object; EIC: Early Immediate Constituents; GEN: genitive; IC: immediate constituent; IO: indirect object;

L: lexical item; LD: lexical domain; MaOP: Maximize Online Processing; MiD: Minimize Domains; MiF: Minimize Forms; N: noun; Nom: Nominative case; NP: noun phrase; NPo: NP with accusative *-o* case particle; OBL: oblique phrase; OS: object before subject; P: preposition or postposition; PCD: phrasal combination domain; Pd: a PP interdependent with V; PGCH: Performance-Grammar Correspondence Hypothesis; Pi: a PP independent of V; Plur: plural; Possp: possessive phrase; PP: prepositional or postpositional phrase; PPL: longer PP; PPm: postpositional phrase; PPS: shorter PP; Rel: relative clause; S: sentence or clause; S': clause with one bar level; Sing: singular; SO: subject before object; SOV: subject-object-verb; SU: subject; UG: Universal Grammar; V: verb; VP: verb phrase.

2 Notice that sequences of [V PP PP] in English are compatible with different attachment options for the second PP. It could be attached to an NP within the first PP, to the VP, or to a higher S or IP. Low attachment to NP within the first PP will generally rule out the permutation option, and the predictions made here for relative ordering do not differ substantially between VP and S attachments (cf. Hawkins, 1994). There are multiple factors that can impact attachment preferences in on-line processing (plausibility, preceding context, frequency, etc.), as MacDonald, Pearlmutter, and Seidenberg (1994) have shown, and the calculation of domain sizes in (3) is made in effect from the speaker's perspective. The speaker knows that the second PP is permutable and is not to be attached within the first. Even from the hearer's perspective, however, notice that the second PP is not reached until the ninth word after the verb in (3b) compared with the the fifth word in (3a), and hence in all the structures in which high attachment is assumed on-line, (3b) will be a less minimal processing domain for phrase structure assignments than (3a).

3 Gibson's (1998) "locality" principle makes many similar predictions, and the wealth of experimental evidence that he summarizes there supports the MiD principle here.

4 In the parsing theory of Hawkins (1990, 1993, 1994), demonstrative determiners can construct NP just as N can, and this may explain the equal productivity of DemN and NDem in head-initial languages; both orders can construct NP at its outset.

5 The prediction I make for Cantonese performance is that the resumptive pronoun should be preferred in proportion to the complexity of phrases that intervene between the subcategorizer in the relative and the head of the relative. Cf. Hawkins (2004, p. 175) for the definition of a "filler-gap domain" that is assumed here (roughly the smallest domain linking the head to the position relativized on, either a gap or a subcategorizor).

References

Aissen, J. (1999). Markedness and subject choice in Optimality Theory. *Natural Language & Linguistic Theory*, *17*, 673–711.

Aldai, G. (2003). The prenominal [-Case] relativization strategy of Basque: Conventionalization, processing, and frame semantics, USC, UCLA: MS, Depts. of Linguistics.

Allwood, J. (1982). The complex NP constraint in Swedish. In E. Engdahl & E. Ejerhed (Eds.), *Readings on unbounded dependencies in Scandinavian languages*. Stockholm: Almqvist & Wiksell.

Ariel, M. (1999). Cognitive universals and linguistic conventions: The case of resumptive pronouns. *Studies in Language*, *23*, 217–269.

Bowerman, M. (1988). The "No Negative Evidence" problem: How do children avoid constructing an overly general grammar? In J. A. Hawkins (Ed.), *Explaining language universals*. Oxford: Blackwell.

Bresnan, J., Dingare, S., & Manning, C. D. (2001). Soft constraints mirror hard constraints: Voice and person in english and lummi. In M. Butt & T. H. King (Eds.), *Proceedings of the LFG 01 conference*. Stanford: CSLI Publications.

Bybee, J., & Hopper, P. (Eds.). (2001). *Frequency and the emergence of linguistic structure*. Amsterdam: John Benjamins.

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

Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris.

Chomsky, N. (1995). *The minimalist program*. Cambridge, MA: MIT Press.

Cinque, G. (2005). Deriving Greenberg's universal 20 and its exceptions. *Linguistic Inquiry*, *36*, 315–332.

Clancy, P. M., Lee, H., & Zoh, M. (1986). Processing strategies in the acquisition of relative clauses. *Cognition*, *14*, 225–262.

Cole, P. (1976). An apparent asymmetry in the formation of relative clauses in modern Hebrew. In P. Cole (Ed.), *Studies in modern Hebrew syntax and semantics*. Amsterdam: North Holland.

Comrie, B. (1973). Clause structure and movement constraints in Russian. In C. Corum, T. C. Smith-Stark, & A. Weiser (Eds.), *You take the high road and I'll take the low node* (pp. 291–304). Chicago: Chicago Linguistic Society.

Comrie, B. (1989). *Language universals and linguistic typology* (2nd ed.). Chicago: University of Chicago Press.

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

Culicover, P. W. (1999). *Syntactic nuts*: *Hard cases, syntactic theory, and language acquisition*. Oxford: Oxford University Press.

Diessel, H., & Tomasello, M. (2006). A new look at the acquisition of relative clauses. *Language*, *81*, 882–906.

Dowty, D. R. (1991). Thematic proto-roles and argument selection. *Language*, *75*, 547–619.

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

Elman, J. L., Bates, E., Johnson, M., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness*: *A connectionist perspective on development*. Cambridge, MA.: MIT Press.

Fodor, J. D. (2001). Setting syntactic parameters. In M. Baltin & C. Collins (Eds.), *The handbook of contemporary syntactic theory* (pp. 730–738). Oxford: Blackwell.

Gell-Mann, M. (1992). Complexity and complex adaptive systems. In J. A. Hawkins & M. Gell-Mann (Eds.), *The evolution of human languages*. Redwood City, CA: Addison-Wesley.

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

Givón, T. (1979). *On understanding grammar*. New York: Academic 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*. Cambridge, MA: MIT Press.

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

Greenberg, J. H. (1995). The diachronic typological approach to language. In M. Shibatani & T. Bynon (Eds.), *Approaches to language typology*. Oxford: Clarendon Press.

Haspelmath, M. (1999). Optimality and diachronic adaptation. *Zeitschrift für Sprachwissenschaft*, *18*, 180–205.

Haspelmath, M. (2006). Parametric versus functional explanations of syntactic universals. Leipzig: MS, Max Planck Institute for Evolutionary Anthropology.

Hauser, M., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*, *298*, 1569–1579.

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

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

Hawkins, J. A. (1993). Heads, parsing, and word order universals. In G. G. Corbett, N. M. Fraser, & S. McGlashan (Eds.), *Heads in grammatical theory*. Cambridge: Cambridge University Press, 231–265.

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

Hawkins, J. A. (1999). Processing complexity and filler-gap dependencies across grammars. *Language*, *75*, 244–285.

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. (2001). Why are categories adjacent? *Journal of Linguistics*, *37*, 1–34.

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

Hoekstra, T., & Kooij, J. G. (1988). The innateness hypothesis. In J. A. Hawkins (Ed.), *Explaining language universals*. Oxford: Blackwell.

Jackendoff, R. (1977). *X-bar syntax*: *A study of phrase structure*. Cambridge, MA: MIT Press.

Jaeger, T. F., & Wasow, T. (2005). Production-complexity driven variation: Relativizer omission in non-subject-extracted relative clauses. Paper presented at the 18th Annual CUNY Sentence Processing Conference, Tuscon, Arizona.

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

Keenan, E. L. (1975). Variation in universal grammar. In R. Fasold & R. Shuy (Eds.), *Analyzing variation in English* (pp. 136–148). Washington, DC: Georgetown University Press.

Keenan, E. L. (1979). On surface form and logical form. *Studies in the Linguistic Sciences*, *8*, 163–203.

Keenan, E. L., & Comrie, B. (1977). Noun phrase accessibility and universal grammar. *Linguistic Inquiry*, *8*, 63–99.

Keenan, E. L., & Hawkins, S. (1987). The psychological validity of the accessibility hierarchy. In E. L. Keenan, *Universal grammar*: *15 essays*. London: Croom Helm.

Kempson, R., Meyer-Viol, W., & Gabbay, D. (2001). *Dynamic syntax*. Oxford: Blackwell.

Kirby, S. (1999). *Function, selection, and innateness: The emergence of language universals*. Oxford: Oxford University Press.

Kvam, S. (1983). *Linksverschachtelung im Deutschen und Norwegischen*. Tübingen: Max Niemeyer Verlag.

Lightfoot, D. W. (1991). *How to set parameters*. Cambridge, MA: MIT Press.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*(4), 676–703.

MacDonald, M. C. (1999). Distributional information in language comprehension, production, and acquisition: Three puzzles and a moral. In B. MacWhinney (Ed.), *The emergence of language*. Mahwah, NJ: Erlbaum.

MacWhinney, B. (1982). Basic syntactic processes. In S. Kuczaj (Ed.), *Language acquisition: Syntax and semantics*. Mahwah, NJ: Lawrence Erlbaum.

Manning, C. D. (2003). Probabilistic syntax. In R. Bod, J. Hay & S. Jannedy (Eds.), *Probability theory in linguistics*. Cambridge, MA: MIT Press, 289–341.

Matthews, S., & Yip, V. (2003). Relative clauses in early bilingual development: Transfer and universals. In A. G. Ramat (Ed.), *Typology and second language acquisition*. Berlin: De Gruyter.

Miyamoto, E. T. (2006). Understanding sentences in Japanese bit by bit. *Cognitive Studies: Bulletin of the Japanese Cognitive Science Society*, *13*, 247–260.

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

Newmeyer, F. J. (2005). *Possible and probable languages: A generative perspective on linguistic typology*. Oxford: Oxford University Press.

Phillips, C. (1996). Order and structure. Ph.D. dissertation, MIT.

Pollard, C., & Sag, I. A. (1987). *Information-based syntax and semantics, Vo1. 1: Fundamentals*, CSLI Lecture Notes No.13. Stanford: Stanford University.

Pollard, C., & Sag, I. A. (1994). *Head-driven phrase structure grammar*. Chicago: The University of Chicago Press.

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

Quirk, R. (1957). Relative clauses in educated spoken English. *English Studies*, *38*, 97–109.

Reali, F., & Christiansen, M. H. (2007a). Processing of relative clauses is made easier by frequency of occurrence. *Journal of Memory and Language*, *57*, 1–23.

Reali, F., & Christiansen, M. H. (2007b). Word chunk frequencies affect the processing of pronominal object-relatives. *The Quarterly Journal of Experimental Psychology*, *60*, 161-170.

Saah, K. K., & Goodluck, H. (1995). Island effects in parsing and grammar: Evidence from akan. *Linguistic Review*, *12*, 381–409.

Schütze, C. T., & Gibson, E. (1999). Argumenthood and English prepositional phrase attachment. *Journal of Memory and Language*, *40*, 409–431.

Sheldon, A. (1974). On the role of parallel function in the acquisition of relative clauses in English. *Journal of Verbal Learning and Verbal Behavior*, *13*, 272–281.

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

Steedman, M. (2006) A Zipfian view of Greenberg 20. MS, University of Edinburgh.

Svenonius, P. (2000). Introduction. In P. Svenonius (Ed.), *The derivation of VO and OV*. Amsterdam: John Benjamins.

Tomasello, M. (2003). *Constructing a language*. Cambridge, MA: Harvard University Press.

Tomlin, R. S. (1986). *Basic word order*: *Functional principles*. London: Croom Helm.

Wasow, T. (1997). Remarks on grammatical weight. *Language Variation and Change*, *9*, 81–105.

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

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