Title

Typological Variation and Efficient Processing

ISBN

978-1-118-30175-3

Author

Hawkins, John A

Publication Date

2015

Peer reviewed

Typological Variation and Efficient Processing

John A. Hawkins, UC Davis & Cambridge University

X.1 Introduction

O'Grady (2008:448) defines emergentism within the language sciences as follows:

(X.1) The emergentist thesis for language
The phenomena of language are best explained by reference to more basic non-linguistic (i.e. non-grammatical) factors and their interaction – physiology, perception, processing, working memory, pragmatics, social interaction, properties of the input, the learning mechanisms, and so on.

This chapter presents an approach to cross-linguistic variation and typology that is very much in this tradition. I will summarize a research program in which typological patterns are seen as conventionalizations of efficient processing routines, see Hawkins (1990, 1994, 1999, 2004, 2009ab, 2014). The preferences can be observed in performance data from languages that permit structural choices and alternatives, e.g. between competing word orders, or between relative clauses with and without a relativizer or resumptive pronoun. The patterns and principles of performance are the same as those we find in the fixed conventions of languages with fewer options (i.e. with more fixed orderings or gaps only in certain relativization environments), and this leads to a Performance-Grammar Correspondence Hypothesis (PGCH).

I begin by defining this hypothesis and summarizing some converging research from different perspectives that supports it, followed by some brief correspondences between performance and grammars (§X.2). I then give some general principles I have proposed to define efficient processing (Minimize Domains, Minimize Forms, and Maximize On-line Processing), with illustrative performance and cross-linguistic data (§X.3). The next section (§X.4) discusses the mechanisms of emergence: how exactly do conventions emerge diachronically out of performance preferences, and in what sociolinguistic and learning environments? Finally (in §X.5) I discuss how conventionalization impacts efficient processing and I raise the question of whether the conventions that emerge in different languages are grammatical rules in the traditional sense or fixed processing routines, as proposed by O'Grady (2005).

X.2 The Performance-Grammar Correspondence Hypothesis

In my 2004 book I defined the PGCH as follows:

(X.2) 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.

In Hawkins (1990, 1994, 2004) I argued that the preferred word orders in languages with choices are those that are most productively conventionalized as fixed orders in languages with less freedom. Givón (1979:26-31) was one of the first to observe that performance preferences in one language corresponded to an actual categorical requirement for the relevant rule or property in another. The considerable preference for definite over indefinite grammatical subjects in English, for example, has been conventionalized into a categorical requirement for definite subjects in Krio and other languages. More recently Bresnan et al. (2001) refer to cases of this kind as 'soft constraints' in some languages becoming 'hard constraints' in others. In morpho-syntax Greenberg (1966) drew attention to corresponding patterns between performance and grammars in his discussion of markedness hierarchies like Singular > Plural > Dual > Trial/Paucal. Morphological inventories across grammars and declining allomorphy provided evidence for the universal hierarchies, while declining frequencies of use in languages with rich inventories suggested not only a correlation with performance but a possibly causal role for it in the evolution of the grammatical regularities themselves (Greenberg 1995:163-164).

There is now a growing awareness of this performance-grammar correspondence in many branches of the language sciences. Haspelmath (1999) has proposed a theory of diachrony in which usage preferences lead to changing grammatical conventions over time. Bybee & 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) were incorporated in the simulation and led 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 et al. 2001, Manning 2003) incorporates both the 'soft constraints' and the 'hard constraints' referred to above. Newmeyer (2005) advocates replacing generative parameters with principles derived from language processing, while Phillips (1996) and Kempson et al. (2001) incorporate the on-line processing of language into the rules and representations of the grammar.

This is all a far cry from the kind of performance-grammar relationship originally advocated in Chomsky (1965:11-12). Although the (competence) grammar was an important component of an overall performance model in his theory, that had to be constantly accessed in language use, he argued that performance factors including memory limitations had given nothing back to grammars. Instead these were claimed to be autonomous and independent of performance and determined ultimately by an innate U(niversal) G(rammar). In more recent work (e.g. 1995, 2005) Chomsky does not seem to have changed his basic view that grammars are immune to performance, despite appeals to so-called 'third factors' (see Mobbs 2008 for extensive discussion and Hawkins 2014:ch.3.6 for a summary and critique).

The methodology that my collaborators and I have been using in order to try and clarify this issue has been an empirical and interdisciplinary one. We have been systematically comparing variation patterns within and across languages, i.e. in usage and in grammars. The research has proceeded as follows. First, find a language whose grammar generates or permits a plurality of structural alternatives of a common type and examine their distribution in language use. They may involve alternative orderings of the same constituents with the same or similar domination relations in the phrase structure tree, e.g. different orderings of NP and PP in the post-verbal domain of Hungarian with its freely ordered constituents (Kiss 2002), or [PP NP V]vp vs. [NP PP V]vp in a verb-final language like Japanese (Kuno 1973). Or they may involve alternative relative clauses with and without an explicit relativizer, as in English (the Danes whom/that he taught vs. the Danes he taught) (see Wasow et al. 2011), or alternations between relativizations on a direct object using a gap strategy vs. a resumptive pronoun strategy (in sentences corresponding to the Danes that he taught versus the Danes that he taught them), as in Hebrew (Ariel 1999).

Second, check for the distribution of these same structural patterns in grammars across languages. The PGCH predicts that when the grammar of one language is more restrictive and eliminates one or more structural options that are permitted by the grammar of another, the restriction will be in accordance with performance preferences. The preferred structure will be retained and 'fixed' as a grammatical convention, the dispreferred structures will be removed. Either they will be eliminated altogether from the output of the grammar or they may be retained in some marginal form as lexical exceptions or as limited construction types. Some performance-grammar correspondences that result from this are summarized briefly here.

(i) The Keenan & Comrie (1977) Accessibility Hierarchy (SU>DO>IO/OBL>GEN, see Comrie 1989) has been much discussed in this context. Grammatical cut-off points in relative clause formation possibilities across languages follow the hierarchy, and Keenan & Comrie argued for an explanation in terms of declining ease of processing down the lower positions on the hierarchy. As evidence they pointed to usage data from languages with many relativizable positions, especially English. In such languages corpus frequencies decline down the hierarchy while processing load and working memory demands have been shown to increase under experimental conditions (Keenan 1975, Keenan & S. Hawkins 1987, Hawkins 1999, Diessel & Tomasello 2006, Kwon et al. 2010).

(ii) More generally, filler-gap dependency hierarchies for relativization and Wh-movement across grammars appear to be structured by the increasing complexity of the permitted gap environments. The 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), see Hawkins (1999, 2004:ch.7).

(iii) Reverse hierarchies across languages for conventionalized gaps in simpler relativization domains and resumptive pronouns in more complex environments 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, Matthews & Yip 2003, Hawkins 2004).

(iv) Parallel function effects (whereby the head of the relative matches the position relativized on) have been shown to facilitate relative clause processing and acquisition

(Sheldon 1974, MacWhinney 1982, Clancy et al. 1986). They also extend relativization possibilities beyond normal constraints holding in languages such as Basque and Hebrew (Aldai 2003, Cole 1976, Hawkins 2013).

(v) 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 1994:315-321).

(vi) (Nominative) subject before (accusative) object 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 (Hawkins 1994, 2013, Gibson 1998, Tomlin 1986, Primus 1999, Miyamoto 2006).

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

(viii) Performance preferences in favor of a definite rather than an indefinite grammatical subject, e.g. in English, correspond to a categorical requirement for a definite subject in others (e.g. in Krio, Givón 1979).

(ix) Performance preferences for subjects that obey the Person Hierarchy (1st,2nd > 3rd) 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 et al. 2001). Sentences corresponding to The boy hit me are ungrammatical in Lummi.

(x) The distinction between zero agreement in local NP environments versus explicit agreement non-locally in the grammar of Warlpiri matches the environments in which zero and explicit forms are preferred in performance in languages with choices, for example in the distribution of zero and explicit relativizers in English (Hawkins 2004:160).

These ten examples are just the tip of a large iceberg of performance-motivated cross-linguistic patterns. They provide strong support for O'Grady's emergentist thesis presented in (X.1). What needs to be specified now is: what is it exactly about the preferred selections in performance and grammars that makes them more efficient? In Hawkins (2004, 2009ab, 2014) I have defined three general principles which I shall summarize briefly here.

X.3 Efficiency Principles

X.3.1 Principle 1: Minimize Domains

Efficiency is increased, first, by minimizing the domains (i.e. the sequences of linguistic forms and their conventionally associated properties) within which certain properties are assigned. This reduces the time course and the processing effort required for the assignment of these properties. This principle is defined in (X.3) (see Hawkins 2004:31):

(X.3) Minimize Domains (MiD)
The human processor prefers to minimize the connected sequences of linguistic forms and their conventionally associated syntactic and semantic properties in which relations of combination and/or dependency are processed. The degree of this preference is proportional to the number of relations whose domains can be minimized

in competing sequences or structures, and to the extent of the minimization difference in each domain.

This principle is a more general version of the Early Immediate Constituents (EIC) principle of Hawkins (1994) which applied only to word order and the parsing of phrase structure. MiD now claims that all syntactic and semantic properties that are assigned in the comprehension and production of sentences are preferably processed within minimal domains.

Consider in this context just the relative ordering of words and phrases. These have to be assembled into the kinds of phrasal 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 a mother phrase M and its immediate constituent daughters (ICs), making phrasal combination faster. Compare (X.4a) and (b):

(X.4) a. The boy vp[looked pp1[for his father] pp2[in the dark and very frightening tunnel]]
1 2 3 4 5
---

b. The boy vp[looked pp2[in the dark and very frightening tunnel] pp1[for his father]]
1 2 3 4 5 6 7 8 9

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

(X.5) Phrasal Combination Domain (PCD) [Hawkins 2004:107]
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.

(X.6) Early Immediate Constituents (EIC) [Hawkins 1994:69-83]
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 languages like English, e.g. for short before long PPs in (X.4). 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 (X.4a) is 3/5 or 60% (5 words required for the recognition of 3 ICs). The comparable ratio for (X.4b) is 3/9 or 33% (9

words required for the same 3 ICs). (For comparable benefits within a Production Model, see Hawkins 2004:106).

Structures like (X.4) were selected from a corpus on the basis of a permutation test (Hawkins 2000): 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 was (PPS = shorter PP, PPL = longer PP):

(X.7) 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)

For a head-final language like Japanese, on the other hand, the categories that construct mother nodes (V, P, Comp, case particles, etc.) are on the right of their respective sisters, and their mothers (VP, PP, …) will be constructed on the right edge in parsing unless there is some alternative constructor that precedes the head (see Hawkins 1994:ch.6, 2004:ch.5, 2014:ch.5 for extensive discussion). This means that long before short orders now provide minimal PCDs in structures containing several phrases, i.e. the mirror-image of English. For example, if the direct object of a verb is a complement clause headed by to, as in (X.8), the distance between this complementizer and the other constituents of the matrix clause in (X.8b), the subject Mary ga and the verb it-ta, is short, just as short in fact as it is in the mirror-image English translation Mary said that ... . Hence the Phrasal Combination Domain for the matrix clause in (X.8b), with its long initial complement clause preceding shorter phrases and categtories, is minimal. In (X.8a), by contrast, with a center-embedded complement clause following a shorter subject NP, the PCD for the matrix clause proceeds all the way from Mary ga to it-ta, and is much longer.

(X.8) a. Mary ga [[kinoo John ga kekkonsi-ta to]s it-ta]vp (Japanese)
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

For similar reasons a long-before-short preference is predicted for [{NPo, PPm} V] structures in Japanese, in alternations such as (X.9) (with -o standing for the accusative case particle, and PPm for a postpositional phrase with a head-final postposition):

(X.9) a. (Tanaka ga) [[Hanako kara]pp [sono hon o]np katta]vp (Japanese)
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

Corpus and experimental support for this preference in Japanese are given in Hawkins (1994:152) (based on data supplied by Kaoru Horie) and in Yamashita (2002) and Yamashita & Chang (2001, 2006), where it is shown that the longer IC is increasingly preferred to the

left in the Japanese clause, whereas it is increasingly preferred to the right in English (see (X.7)). This underscores an important principle for psycholinguistic models: the directionality of weight effects depends on the language type.

A possible explanation for these minimal domains in English and Japanese can be given in terms of reduced processing demands in working memory. If, in (X.4a), the same phrase structure information can be derived from a 5-word viewing window as from 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 (see Lewis & Vasishth 2005). Overall fewer demands will be made on working memory and on the computational system. (X.4a) is more efficient, therefore, but not because some claimed capacity limit has been breached. All the attested orderings in (X.7) 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 MiD predicts that all syntactic and semantic relations are preferably processed in minimal domains. More generally still, Hawkins (2009b) and (2014:ch.2.4) appeals to efficiency in place of constrained capacity in working memory as the explanation for these structural selections in performance and defines efficiency as follows: Communication is efficient when the message intended by speaker S is delivered to hearer H in rapid time and with the most minimal processing effort that can achieve this communicative goal.

Conventionalized orderings in grammars reveal the same degrees of preference for minimal domains and the same mirror-image weight effects that we see in performance. For example, Greenberg (1963) examined alternative verb positions across languages and their correlations with prepositions and postpositions in phrases corresponding to (X.10):

(X.10) a. vp[went pp[to the store]]
---

b. [[the store to]pp went]vp
---

c. vp[went [the store to]pp]
---

d. [pp[to the store] went]vp

(X.10a) is the English order, (X.10b) is the Japanese order, and these two sequences, with adjacent lexical heads (V and P), are massively preferred in cross-linguistic samples, over the inconsistently ordered heads in (X.10c) and (X.10d). (X.11) summarizes the distribution using the database of Dryer's (1992) paper on the 'Greenbergian correlations' (see Hawkins 2004:124):

(X.11) a. vp[V pp[P NP]] = 161 (41%)
IC-to-word: 2/2 = 100%

b. [[NP P]pp V]vp = 204 (52%)
IC-to-word: 2/2 = 100%

c. vp[V [NP P]pp] = 18 (5%)
IC-to-word: 2/4 = 50%

d. [pp[P NP] V]vp = 6 (2%)
IC-to-word: 2/4 = 50%

Assume: V = 1 word; P = 1; NP = 2
EIC-preferred (X.11a)+(b) = 365/389 (94%)

The adjacency of V and P and consistent head ordering to the left or right guarantees the smallest possible string of words (indicated by the underlinings in (X.10)) and the highest IC-to-word ratios for the recognition and construction of VP and its two immediate constituents (ICs), namely V and PP. Non-adjacent V and P in (X.11cd) require longer and less efficient strings for the parsing of phrase structure. I.e. positioning heads next to one another provides a minimal phrasal combination domain for the construction of VP and its daughters. Similarly, in the performance data of (X.7) 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 (X.3), therefore. Conventions of ordering have emerged out of performance preferences, and one and the same principle explains both the preferred structural selections in performance (in languages and structures in which speakers have a choice) and the preferred conventions of grammar. MiD can also explain why there are two productive mirror-image types here, head-initial and head-final, exemplified by (X.10a) and (b): they are equally good strategies for phrase structure comprehension and production (Hawkins 2004:123-6). An efficiency approach can also explain exceptions to the majority patterns and to grammatical principles such as consistent head ordering, as well as many other conventionalized grammatical patterns (see Hawkins 2004, 2014). This approach can motivate and explain what is at best stipulated in alternative accounts based on innate parameters (see Newmeyer 2005).

X.3.2 Principle 2: Minimize Forms

A second general efficiency factor defended at length in Hawkins (2004, 2014) is Minimize Forms:

(X.12) 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.

The basic premise of MiF is that the processing of linguistic forms and of conventionalized property assignments requires effort. Minimizing forms and form-property pairings reduces that effort by fine-tuning it to the information that is already active in communication through accessibility, inferencing, and frequency. This principle is visible in two complementary sets of cross-linguistic and intralinguistic variation data. One set involves preferences for minimal expression in phonology, morphology and syntax (e.g. zero forms) in proportion to ease of processing, including frequency. Nominative case is more frequent than accusative, and singular than plural, etc. Correspondingly, nominative and singular are more often expressed by zero forms than accusative and plural respectively (compare singular dog in English with plural dogs). Another data set involves the number and nature of lexical and grammatical distinctions that languages conventionalize. The preferences are again in proportion to ease of processing, including frequency in performance.

There are countless examples showing that a reduction in form processing is efficient, as long as relevant information can be recovered. Consider the use of pronouns versus full NPs (he/she versus the teacher). The high versus low accessibility of entities referred to in discourse correlates with less versus more formal structure respectively (see Ariel 1990). Also relevant here are Zipf (1949) effects (e.g. the shorter gas replaces gasoline in this high-frequency word), compounds (paper plate is reduced from plate made of paper; paper factory from factory that makes paper, see Sperber & Wilson's 1995 theory of relevant real-world knowledge activated in the processing of these minimal structures), co-ordinate deletions (John bought 0 and Fred ate the pizza, see Hawkins 2004: 93-95 and van Oirsouw 1987), and control structures involving understood subjects of verbs within non-finite subordinate clauses (whose controllers are in a structurally accessible matrix clause position, see Hawkins 2004: 97-101). Filler-gap dependencies in relative clauses (the professor*i* that I know O*i*) are also plausibly motivated by (X.12). Gaps can be identified by reference to the filler with which they are co-indexed. The result is a more minimal structure than the resumptive pronoun counterparts (the professor*i* that I know him*i*) in languages such as Hebrew. The advantage of minimization disappears, however, in more complex environments in which processing demands and processing domains become larger (Hawkins 2004:182-186, Ariel 1999).

Form reduction is further supported by the Economy Principle of Haiman (1983) and by the data he summarizes from numerous languages. It is also reminiscent of Grice's (1975) second Quantity maxim for pragmatic inferencing ('Do not make your contribution more informative than is required'), and more specifically of Levinson's (2000) Minimization principle derived from it ('Say as little as necessary', that is, produce the minimal linguistic information sufficient to achieve your communicational ends).

Principle (X.12) adds a second factor to this efficiency logic, beyond the simplicity or complexity of the surface forms themselves, and is defined in terms of properties that are conventionally associated with forms. There are numerous semantic and syntactic properties that are frequently occurring across languages and that have priority in grammatical and lexical conventions. The property of causation is often invoked in everyday language use and is regularly conventionalized in the morphology, syntax or lexicon (Comrie 1989, Shibatani 1976). Agenthood and patienthood are also frequently expressed and given systematic (albeit partially different) formal expression in ergative-absolutive, nominative-accusative and active languages (Primus 1999). The very frequent speech acts (asserting, commanding and questioning) are each given distinct formal expression across grammars, whereas less frequent speech acts such as bequeathing or baptizing are assigned separate lexical items, but not a uniquely distinctive construction in the syntax (Sadock and Zwicky 1985). Within the lexicon the property associated with builder is frequently used in performance, that of builder who does rapid but high-quality work much less so. The event of X striking Y is frequently selected, that of X striking Y with X's right hand less so. The more frequently selected properties are conventionalized in single lexemes or unique categories and constructions in all these examples. Less frequently used properties must then be expressed through word and phrase combinations and their meanings must be derived by semantic composition. This makes the expression of more frequently used meanings shorter, that of less frequently used meanings longer, and this pattern matches the first pattern of less versus more complexity in the surface forms themselves correlating with relative frequency. Both patterns make

utterances shorter and the communication of meanings more efficient overall, which is why I have chosen to capture them both in one common Minimize Forms principle (X.12).

MiF makes the following predictions for grammars and performance that are tested extensively in Hawkins (2004, 2014), to which the reader is referred for further details and exemplification:

(X.13) Form Minimization Predictions
- a The formal complexity of each F is reduced in proportion to the frequency of that F and/or the processing ease of assigning a given P to a reduced F (e.g. to zero).
- b The number of unique F:P1 pairings in a language is reduced by grammaticalizing or lexicalizing a given F:P1 in proportion to the frequency and preferred expressiveness of that P1 in performance.

X.3.3 Principle 3: Maximize On-line Processing

A third general efficiency principle involves the timing with which linguistic properties are introduced in on-line processing. There is a clear preference for selecting and arranging linguistic forms so as to provide the earliest possible access to as much of the ultimate syntactic and semantic representation as possible. This preference, called Maximize On-line Processing, also results in a preference for error-free on-line processing since errors delay the assignment of intended properties and increase processing effort. Clear examples can be seen across languages when certain common categories {A, B} are ordered asymmetrically A + B, regardless of the language type, in contrast to symmetries in which both orders are productive [A+B/B+A], e.g. Verb+Object [VO] and Object+Verb [OV]. Some examples of asymmetries are summarized in (X.14) (see Hawkins 2002, 2004:ch.8):

(X.14) Some Asymmetries

(i) Displaced WH preposed to the left of its (gap-containing) clause [almost exceptionless; Hawkins 2002,2004]
Who*i* [did you say O*i* came to the party]

(ii) Head Noun (Filler) to the left of its (gap-containing) Relative Clause
E.g. *the students*i [*that I teach O*i]
If a lg has basic VO, then NRel [exceptions = rare] (Hawkins 1983)

VO	OV
NRel (English)	NRel (Persian)
*RelN	RelN (Japanese)

(iii) Antecedent precedes Anaphor [highly preferred cross-linguistically]
E.g. ***Mary*** *washed* ***herself*** (SVO), *Washed* ***Mary herself*** (VSO), ***Mary herself*** *washed* (SOV) = highly preferred over e.g. *Washed* ***herself Mary*** (VOS)

(iv) Wide Scope Quantifier/Operator precedes Narrow Scope Q/O [preferred]
E.g. ***Every student a book*** *read* (SOV lgs) ∀∃ preferred
A book every student *read* (SOV lgs) ∃∀ preferred

In these examples there is an asymmetric dependency of B on A: the gap is dependent on the head-noun filler in (ii) (for gap-filling), the anaphor on its antecedent in (iii) (for co-indexation), the narrow scope quantifier on the wide scope quantifer in (iv) (the number of books read depends on the quantifier in the subject NP in Every student read a book/Many students read a book/Three students read a book, etc). The assignment of dependent properties to B is more efficient when A precedes, since these properties can be assigned to B immediately in on-line processing. In the reverse B + A there will be delays in property assignments on-line ("unassignments") or misanalyses ("misassignments"). If the relative clause precedes the head noun the gap is not immediately recognized and there are delays in argument structure assignment within the relative clause; if a narrow scope quantifier precedes a wide scope quantifier, a wide scope interpretation will generally be (mis)assigned on-line to the narrow scope quantifier; and so on.

Maximize On-line Processing is defined in (X.15):

(X.15) 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.

These three principles, MiD (X.3), MiF (X.12) and MaOP (X.15) sometimes cooperate and reinforce each other's predicted effects, but sometimes they compete. For example, MaOP defines a filler before gap preference for nominal heads before relative clauses, and this preference is visible in both VO and OV languages (see (X.14ii). Only rigid V-final languages resist this preference, in which there are numerous head-final phrases that prefer to contain a head-final NP (i.e. RelN) in accordance with MiD which competes with MaOP in these structures (see Hawkins 2004:205-209). Cooperation and competition between all of these principles are discussed extensively in Hawkins (2009b) and (2014:ch.9), to which the reader is referred for further examples.

There are other recent approaches to defining efficiency in psycholinguistics which are very much in the spirit of the research program described here and which add useful dimensions to it. For example, sophisticated measures have been developed to test the idea that expectedness and predictability are reflected in ease of processing on a word-by-word basis in a sentence (see Jurafsky 1996, Hale 2003, Levy 2008). It has been proposed that efficiency is achieved through uniform information density over the linguistic signal and over time (Jaeger 2006). These proposals, which are empirically well-supported, provide an additional on-line component to the measurements of efficiency given here and defined on individual words and whole structures (see Jaeger & Tily 2011 for a summary).

X.4 Mechanisms of Emergence

Having set out the evidence for a large number of correspondences between performance preferences and grammars and for the kinds of efficiencies that underlie them, we can now

consider the mechanisms by which conventionalized linguistic knowledge emerges. How exactly do new grammatical forms and rules actually get into the grammar from a preceding stage of the language in which a grammatical convention does not exist? Who innovates them (monolingual or bilingual child learners? second language learners? other sociolinguistic groups?)? And in what linguistic or multi-linguistic environments?

A number of linguists have been concerned with this question and have clarified what Kirby (1999) calls the 'adaptive mechanisms' (by which grammars incorporate the kinds of performance preferences described here), see especially Haspelmath (1999), Croft (1996, 2000), Bybee (1988, 2001), Bybee & Hopper (2001) and Heine (2008). Haspelmath (op. cit.) gives a very Darwinian account of language change and summarizes three crucial ingredients for the evolution of new linguistic conventions. First, there must be variation among structural alternatives at some stage of a language, e.g. alternative orderings of the same phrases (see §X.3.1), or both gaps and resumptive pronouns when relativizing on the same position (see Keenan & Comrie 1977, Hawkins 2004:189, 2014:ch.2.2.3, Ariel 1999), reduced versus less reduced morphological forms (Greenberg 1966, Primus 1999, Hawkins 2004:ch.4), phonological variants, and so on. Second, principles of 'user optimality' apply to these variants at successive historical stages to select some rather than others, resulting in different frequencies of use. In the present context, these principles would include efficiency principles as developed here. And third, the preferred and most frequently used variants of performance may become obligatory and categorical, with less frequent alternatives being lost altogether.

Haspelmath chooses the constraints of Prince & Smolensky's (1993) Optimality Theory as the conventions of grammar that have arisen in response to user optimality in order to illustrate these points, and he discusses how these constraints could evolve from one generation of learners and users to the next. In this context I shall mention some further points involving the role of efficiency in this process.

The relative sequencing of changes, and the innovation of new grammatical variants proceeds gradually, as historical linguistics textbooks point out (see e.g. Campbell 2004 and McMahon 1994). Languages change the ordering of single-word adjectives before they reorder adjective phrases, i.e. they develop variant single-word orders first before phrasal variants. This is the clear diachronic progression that underlies the synchronic typological pattern documented in Dryer (1992), whereby single-word modifiers of heads are often typologically inconsistent with the predominant head ordering of the grammar, as in English yellow book, at the same time that phrasal modifiers are typologically consistent, as in book yellow with age in which the adjective phrase (AdjP) follows the head-initial N. A change involving these word orders goes through a 'word order doubling' stage (Hawkins 1983) in which both orders (e.g. AdjN and NAdj) exist at the same time, as in the Romance languages.

In relativization hierarchies like the Keenan-Comrie (1977) Accessibility Hierarchy, gaps or resumptive pronouns are similarly innovated as competing variants or doublets. They also occur at points of transition between these strategies on that hierarchy (see Hawkins 2004:189). The gap/pro alternations are transitional between higher gap-only positions and lower pro-only positions in the relevant languages. Competing gaps and pronouns do not arise at points that are distant from either strategy. For example, the SU gap is extended first to the DO and only later to the IO or OBL.

From the perspective of the PGCH (X.2) this means that small conventionalized processing load departures from the current grammar precede bigger ones. If you are going to

change word order, change the single-word categories first that I have argued have minimal impact on phrasal combination domains and on the processing of phrase structure (see Hawkins 2004:124-126, 2014:ch.5.5) before changing the phrasal categories that have a bigger impact. Languages, dialects and sociolects innovate certain grammatical changes before others and we accordingly see variation patterns arising in different social and regional speech communities in certain parts of the grammar first and not in others.

The ultimate triggers or causes for these changes can be various, as historical textbooks again point out (Campbell 2004, McMahon 1994). Historical linguists are increasingly coming to recognize the profound role of contact between languages (see. Thomason & Kaufman 1988, Thomason 2001) and of different types of bilingualism that result in transfers from one language to another (see Hawkins & Filipovic 2012 for discussion of transfers from different L1s into L2 English). Trudgill (2011) makes a distinction between contact situations in which there are many non-native adult learners of the L2 versus those with genuinely bilingual native speakers of the two languages. The contact-induced changes resulting from the former, he argues, are different from those in the latter and they include morphological simplifications as a consequence of imperfect learning. Native-speaking bilinguals, on the other hand,, can serve as a conduit for the transfer of more complex linguistic features between languages.

It may not be the bilingual child learners who innovate grammatical changes, therefore, but rather adolescent bilinguals or other social groups. Trudgill's (op. cit.) perspective provides a welcome antidote to what has become almost a dogma in linguistics, to the effect that it is necessarily children who innovate language change (cf. e.g. King 1969, Lightfoot 1991, Kirby 1999). First of all, to the extent that learners do change certain features of the language they are learning that subsequently spread to the rest of the speech community, they may be second language learners who are present in sufficient numbers to have an impact. Second, the language changers may not be children at all (see further the work of Labov 1972, 1994, 2007 for innovations in language among speakers of different ages and social groups).

Contact and bilingualism provide the social conditions for the spread of head-initial and head-final word orders in strikingly contiguous areas of the globe, as seen in Dryer's (2005ab) WALS maps involving VO versus OV and prepositions versus postpositions respectively. Similar basic orders such as these are found adjacent to one another. This explains why the only Austronesian languages to abandon their VO syntax and change to OV are those in New Guinea in contact with indigenous OV languages. See the discussion in M.D. Ross (1996, 2001) of a particularly extreme case of this involving the development of OV patterns in the Western Oceanic language Takia under the influence of Papuan Waskia on the north coast of Papua New Guinea. In Mesoamerica Gast (2007) points to the reverse shift of OV to VO among Uto-Aztecan languages like Nahuatl in contact with VO Mesoamerican languages. Heine (2008) gives an insightful discussion of how new word orders arise diachronically in these kinds of contact situations by building on and extending minority orders that existed before and by adapting already existing structures in the recipient language. In addition to the important role played by non-native speaking adults versus genuine bilinguals in these contact situations Trudgill (2009, 2011) and Lupyan & Dale (2010) provide evidence for further social correlations with structural simplicity and complexity in grammars, such as community size.

Processing efficiency adds three new dimensions to these other, primarily external, causes of change. First, it modulates their sequencing and relative timing: smaller changes in processing load precede larger ones. This constrains the corresponding types of synchronic variation that we find, e.g. word order doublets for single-word modifiers of heads in the absence of doublets for phrasal modifiers, and so on. Second, processing can cause internal adjustments and a ripple effect throughout the grammar in the event that e.g. a head-final OV language changes its verb position to VO. Other phrases will then change their head ordering in response to ease and efficiency of use as domains for phrase structure processing are gradually minimized throughout the whole grammar (see MiD (X.3) above) with smaller adjustments again preceding larger ones. Third, different efficiency principles sometimes cooperate and sometimes compete within languages and structures, as in the relative clause strategies of OV languages (see §X.3.3 above and Hawkins 2004:205-210, 2014:chs.7.3-7.4). It is not possible for each language type to satisfy all the preferences all the time, and this is the essential reason, I believe, why we see some of the variation that we do. The most common language types appear to be good for most principles, albeit not for all, and the efficiencies that cannot be satisfied in a given language type create an internal tension for change, and a readiness for change in the event that external factors such as language contact and language transfer come to favor it (see the next section and Hawkins 2014:ch.9.3).

Processing efficiency therefore helps us understand which structures will change, when they will change, and the directionality of the changes, i.e. what they will change into. Efficiency is a hitherto relatively neglected cause of language change. The variation patterns we see across languages have ultimately arisen, I believe, not because of some mechanism internal to the grammar, nor as a consequence of innate parameters triggered by the data of experience, and not necessarily because of changes introduced by child language learners, but as a result of processing efficiency acting within various social groups of language users to select and conventionalize certain variants from among competing options while gradually eliminating others. Efficiency can be added to the language-external and -internal factors that are routinely recognized and investigated in historical linguistics and sociolinguistics, to give us a more complete model of language change, variation and mixing.

X.5 Conventionalization and Efficiency

An important issue has been raised by O'Grady (2005, 2008). If we accept the emergentist thesis (X.1), and the validity of the kinds of performance-grammar correspondences seen here, do we then need to assume that there are grammatical rules at all in the traditional sense? What exactly is it that gets conventionalized? Rules and principles of a grammar, or certain procedures for processing the forms of a language, that can differ in their formulations just as grammatical rules do, for example when describing the different directionalities of heads and their modifiers in head-initial and head-final languages? O'Grady has argued that grammars as such are not needed and are subsumed by the procedures and fixed routines of an efficient processor.

In the research program described here I have assumed that that there is still a conventionalized competence grammar in the traditional sense (following Chomsky 1965), separate from performance procedures, although I have disagreed with Chomsky on the relationship between performance and this competence grammar (see X.2 above). Grammars

are constantly accessed in performance and, as I see it, have been profoundly shaped by performance pressures as a consequence. But is there then any need to assume a grammar at all separate from an efficient processor that has become set in its ways by developing a set of processing routines of varying strengths? O'Grady answers "no". This reduction leads, prima facie, to a simpler theory and it is attractive for this reason: the processor cannot be eliminated and is needed independently to account for how language is used, but a grammar can in principle be dispensed with in the event that it adds nothing to the operations specified as part of an (efficient) processor.

I see two relevant issues here, which are connected but which do need to be separated. First, are there traditional rules of grammar in addition to processing procedures? Second, how has conventionalization, i.e. the fixing of certain linguistic behavior patterns within a speech community, impacted these rules and/or procedures?

In my work I have focused on, and tried to better understand, the relationship between performance data that have not been fixed by convention (e.g. free word orders in languages and structures that have them), and the conventions of grammars in languages with fewer options (e.g. fixed ordering rules). There is no question that these different conventions are reflected in different procedures for processing languages. But one reason for keeping the traditional idea of a grammar, it has seemed to me, is that the grammar provides a convenient locus in which competing processing demands, as well as the all-important needs of expressive power and meaning, can be reconciled and their resolutions described. The grammar can also account for certain inefficiencies in performance that arise precisely as a consequence of the conventionalization of efficiency in the normal case.

Different conventions across languages are not always efficient in all respects and in every performance instance, even when they can be argued to be motivated by efficient processing in general. Rather, they are as efficient as they can possibly be, given competing demands, and they are efficient in general. This is one reason why I have maintained the traditional notion of a competence grammar, even while arguing, as O'Grady does, for an efficient processor and for efficient structural conventions shaped by it across languages. I have also argued that different language types have conventionalized structures in proportion to their degrees of efficiency, as measured for example by distributional frequencies in the World Atlas of Language Structures database (Haspelmath et al. 2005).

The challenge for O'Grady's theory eliminating grammatical rules altogether lies, as I see it, in whether grammatical conventions can indeed be convincingly subsumed under, and replaced by, the fixed routines of an efficient processor. I am sympathetic to this possibility, as I mentioned above. What needs to be clarified is whether his theory is more than a notational variant of the kind of theory advocated here in which grammars are profoundly shaped by performance but nonetheless retain a certain autonomy as conventionalized grammatical knowledge distinct from the fixed routines of an efficient processor. There are profound philosophical and methodological issues raised by this, which I shall not attempt to resolve here. The important point is that both approaches, O'Grady's and mine, are built on the shared assumption that grammatical properties and cross-linguistic variation are strongly shaped by processing and performance. What I shall illustrate in this final section is how conventionalization of grammatical knowledge can respond in possibly different ways to efficient processing, resulting in typological variation.

First, I have argued that head-initial and head-final languages can each be highly efficient with respect to the operations that combine phrases together (see the kinds of language frequencies summarized in (X.11) supporting Minimize Domains (X.3) and EIC (X.6)), and that this explains their almost equal productivity. But in other respects there are competing and complementary efficiencies between them. A language with verb before object order, like English, will activate possible co-occurrences of the verb at the verb, based on the predictability of a finite number of continuations, prior to their actual occurrence. This provides early access to much of a clause's structure and can be argued to be efficient. But there are also disadvantages to having the verb early. These can include numerous temporary ambiguities and garden paths disfavored by MaOP (X.15), as in the italicized portions of (X.16) and (X.17) in which the answer and the book are first (mis)assigned to their immediately preceding verbs and then reassigned as subjects to their following verbs (see Frazier 1985, Hawkins 2014:ch.7.2):

(X.16) Bill knew the answer was correct.
(X.17) While Mary was reading the book fell down.

Verb-final languages avoid these temporary ambiguities and garden paths to a greater extent, by immediately selecting the correct predicate frame for the verb at the verb, based on preceding arguments that have already been encountered (see further Hawkins 1995). But verb-finality in turn has disadvantages and inefficiencies resulting from the late production of this highly predictive and integrative element within the clause (see Hawkins 2014:ch.7 for full details). Neither verb position is optimally efficient, therefore, but each can be as efficient as possible, and different grammars can then adjust to their inherent inefficiencies and compensate for them, e.g. through rich case marking and early on-line thematic role assignments in verb-late languages (Hawkins 2014:ch.7.2). There are numerous competing efficiencies of this sort and grammars conventionalize some at the expense of others (see Hawkins 2014:ch.9). To give a further example, there appears to be no optimal order for relative clauses in OV languages, and this results in considerable variation contrasting with the uniformity of NRel in VO languages (see §X.3.3).

Second, languages can conventionalize different tolerances for inefficiency in the form of different degrees of permitted complexity in processing. We see a nice reflection of this in the conventions for filler-gap dependencies. Even members of the same language family, like Germanic, with only a couple of thousand years of separation in historical time, permit and use structures with different complexities. In most dialects of German, including the standard language, the following relative clause gap environment is ungrammatical. Its English translation, given in the gloss, is not (see Hawkins 2004:194, based on Kvam 1983):

(X.18) *die Person*i* [die*i* du glaubst [dass Johann O*i* gesehen hat]] (German)
the person who you think that John seen has
'the person who you think that John has seen'

Scandinavian languages like Swedish, on the other hand, permit numerous options not found in English, including the relative clause in (X.19) which violates the Complex Noun Phrase Constraint (Ross 1967) of English (see Allwood 1982, Engdahl 1982 and Hawkins 2004:193):

(X.19) ett ben*i* [som*i* jag ser np[en hund s[som gnager på O*i*]]] (Swedish)
a bone which I see a dog which gnaws on
'a bone which I see a dog which is gnawing on'

The conventionalized tolerance level for processing complexity in these relative clauses differs across the different speech communities. The more complex ones permit more expressive power, i.e. more relative clause possibilities, but at the expense of more effortful processing and of some clear inefficiencies. For example, the most efficient parsing routine for the Swedish (X.19) will garden-path the hearer by linking ben first to ser, giving the interpretation 'a bone which I see'. This will then need to be revised to the interpretation in which the dog is the object of seeing, as given in the gloss.

Third, conventionalizing a rule in response to processing efficiency produces structures that are not always efficient on all occasions. For example, there appears to be a convention of English whereby a center-embedded clause (S) is blocked in structures like (X.20), while a center-embedded NP is permitted in this environment, as in (X.21):

(X.20) *Did s[that the boy failed his exam] surprise Mary?
(X.21) Did np[the boy] surprise Mary?

Embedded clauses are typically much longer than NPs in performance, which makes the phrasal combination domain (X.5) for the matrix clause longer and more complex in (X.20) than in (X.21). We can say that the Minimize Domains preference of (X.3) has been grammaticalized in the selective blocking of (X.20), and this can be fomulated in a grammatical rule or constraint blocking a center-embedded S node. The important point here is that different phrasal types are available to a grammar, S versus NP, and these can be selected by syntactic rules in particular languages to define as ungrammatical certain sentence types that are generally bad for processing, due to the average weights of the phrases in question in performance.

Notice now the consequence of conventionalizing a wellformedness distinction that allows (X.21) while disallowing (X.20). Syntactic rules and constraints apply to phrases and categories, not to terminal strings. So on those occasions in performance when an NP happens to be longer than an embedded clause normally is, the grammaticality facts remain the same. Compare (X.20) and (X.21) with (X.22):

(X.22) Did np[the failure of the boy to pass his exam] surprise Mary?

The center-embedded subject NP in this sentence is longer and more complex than the center-embedded S in (X.20). Yet the NP of (X.22) is grammatical, just as (X.21) is grammatical, and moreover it provides a structure in which English speakers can express the meaning intended by the ungrammatical (X.20). The grammar is sensitive to the categorial status of these word groupings, not to their length on particular occasions. I.e. because the length and complexity of NP is typically less than that of an embedded clause, English has grammaticalized a distinction between them which disallows (X.20), but permits (X.21) to be generated as grammatical. As a result the long center-embedded NP of (X.22) can survive as

grammatical to express the meaning of (X.20). The typical and average weights of these phrases are arguably what led the grammar to make the distinction between center-embedded S and NP in the first place. The resulting grammar then generates some inefficient structures which it cannot block because of the way that grammars conventionalize linguistic rules, even those that are designed to generate efficient structures in most cases.

A grammar with conventionalized rules that is accessed in language use by production mechanisms gives us a way of explaining why structures can be produced that are sometimes inefficient, even when the production system is normally highly efficient, and even when the grammar has plausibly conventionalized the formal rule in question in response to performance frequencies and ease of processing. A grammar separate from a processor provides a ready way of explaining these kinds of facts. It remains to be seen whether these competing demands on a grammar and competing efficiencies in processing can be captured and explained within an emergentist theory consisting only of an efficient processor and without a separate grammar.

Regardless of the outcome of this more technical debate, I see the research program presented here as offering strong support for the emergentist thesis of O'Grady (2005, 2008) (see (X.1)), in the area of language typology and language change. Typological variants and different structural properties have emerged that are generally as efficient as they can possibly be for processing, and the relative frequencies with which they have emerged are in proportion to degrees of efficiency of the kind defined in this chapter in §§X.3.1-X.3.3.

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 rtrategy of Basque: Conventionalization, processing and frame semantics', MS, Departments of Linguistics, USC and UCLA.

Allwood, J. (1982) 'The Complex NP Constraint in Swedish', in E. Engdahl and E. Ejerhed (eds.).

Ariel, M. (1990) Accessing Noun-Phrase Antecedents. Routledge, London.

Ariel, M. (1999) 'Cognitive universals and linguistic conventions: The case of resumptive Pronouns', Studies in Language 23:217-269.

Bresnan, J., Dingare, S. and Manning, C.D. (2001) 'Soft constraints mirror hard constraints: Voice and person in English and Lummi', in M. Butt and T.H. King (eds.), Proceedings of the LFG 01 Conference, CSLI Publications, Stanford.

Bybee, J. (1988) 'The diachronic dimension in explanation', in J.A. Hawkins (ed.), Explaining Language Universals, Blackwell, Oxford, 350-379.

Bybee, J. (2001) 'Mechanisms of change in grammaticalization: The role of frequency', in R.D. Janda & B.D. Joseph (eds.), Handbook of Historical Linguistics, Blackwell, Oxford, 602-623.

Bybee, J. and Hopper, P. (eds.) (2001) Frequency and the Emergence of Linguistic Structure. John Benjamins, Amsterdam

Campbell, L. (2004) Historical Linguistics: An Introduction, 2nd edn.. MIT Press, Cambridge,

Mass..
Chomsky, N. (1965) Aspects of the Theory of Syntax. MIT Press, Cambridge, Mass..
Chomsky, N. (1995) The Minimalist Program. MIT Press, Cambridge, Mass..
Chomsky, N. (2005) 'Three factors in language design', Linguistic Inquiry 36:1-22.
Clancy, P.M., Lee, H. and 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, North Holland, Amsterdam, 231-247.
Comrie, B. (1989) Language Universals and Linguistic Typology, 2nd edn.. University of Chicago Press, Chicago.
Croft, W. (1996) 'Linguistic selection: An utterance-based evolutionary theory of language change', Nordic Journal of Linguistics 19:99-139.
Croft, W. (2000) Explaining Language Change: An Evolutionary Approach, Pearson Education
Croft, W. (2003) Typology and Universals, 2nd edn.. Cambridge University Press, Cambridge.
Diessel, H., and Tomasello, M. (2006) 'A new look at the acquisition of relative clauses', Language 81: 882-906.
Dryer, M.S. (1992) 'The Greenbergian word order correlations', Language 68: 81-138.
Dryer, M.S. (2005a) 'Order of object and verb', in M. Haspelmath et al. (eds.), 338-341.
Dryer, M.S. (2005b) 'Order of adposition and noun phrase', in M. Haspelmath et al. (eds.), 346-349.
Engdahl, E. (1982), 'Restrictions on unbounded dependencies in Swedish', in E. Engdahl and E. Ejerhed, eds., 151-174.
Engdahl, E. and Ejerhed, E. (1982) Readings on Unbounded Dependencies in Scandinavian Languages, Almqvist & Wiksell, Stockholm.
Frazier, L. (1985) 'Syntactic complexity', in D. Dowty, L. Karttunen and A. Zwicky (eds.), Natural Language Parsing, Cambridge University Press, Cambridge, 129-189.
Gast, V. (2007) 'From phylogenetic diversity to structural homogeneity: On right-branching constituent order in Mesoamerica', Sky Journal of Linguistics 20: 171-202.
Gibson, E. (1998) 'Linguistic complexity: Locality of syntactic dependencies', Cognition 68: 1-76.
Givòn, T. (1979) On Understanding Grammar. Academic Press, New York.
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, MIT Press, Cambridge, Mass., 73-113.
Greenberg, J.H. (1966) Language Universals, with Special Reference to Feature Hierarchies. Mouton, The Hague.
Greenberg, J.H. (1995) 'The diachronic typological approach to language', in M. Shibatani and T. Bynon (eds), Approaches to Language Typology, Clarendon Press, Oxford.
Grice, H.P. (1975) 'Logic and conversation', in P. Cole and J. Morgan (eds.), Speech Acts, Academic Press, New York, 41-58.
Haiman, J. (1983) 'Iconic and economic motivation', Language 59: 781-819.
Hale, J. (2003) 'The information conveyed by words in sentences', Journal of Psycholinguistic Research 32:101-123.

Haspelmath, M. (1999) 'Optimality and diachronic adaptation', Zeitschrift für Sprachwissenschaft 18: 180-205.
Haspelmath, M., Dryer, M.S., Gil, D. and Comrie, B. (eds.) (2005), The World Atlas of Language Structures, Oxford University Press, Oxford.
Hawkins, J.A. (1983), Word Order Universals. Academic Press, New York.
Hawkins, J.A. (1990) 'A parsing theory of word order universals', Linguistic Inquiry 21: 223-261.
Hawkins, J.A. (1994) A Performance Theory of Order and Constituency. Cambridge University Press, Cambridge.
Hawkins, J.A. (1995) 'Argument-predicate structure in grammar and performance: A comparison of English and German', in I. Rauch and G.F. Carr (eds.), Insights in Germanic Linguistics, de Gruyter, Berlin, 127-144.
Hawkins, J.A. (1999) 'Processing complexity and filler-gap dependencies', Language 75: 244-285.
Hawkins, J.A. (2000) 'The relative order of prepositional phrases in English: Going beyond manner-place-time', Language Variation and Change 11: 231-266.
Hawkins, J.A. (2002) 'Symmetries and asymmetries: Their grammar, typology and parsing', Theoretical Linguistics 28.2: 95-149.
Hawkins, J.A. (2004) Efficiency and Complexity in Grammars. Oxford University Press, Oxford.
Hawkins, J.A. (2009a) 'Language universals and the performance-grammar correspondence hypothesis', in M.H. Christiansen, C. Collins & S. Edelman (eds.), Language Universals, Oxford University Press, Oxford, 54-78.
Hawkins, J.A. (2009b) 'An efficiency theory of complexity and related phenomena', in G. Sampson, D. Gil & P. Trudgill (eds.), Language Complexity as an Evolving Variable, OUP, Oxford, 252-268.
Hawkins, J.A. (2014) Cross-linguistic Variation and Efficiency. Oxford University Press, Oxford.
Hawkins, J.A. and Filipović, L. (2012) Criterial Features in L2 English: Specifying the Reference Levels of the Common European Framework. Cambridge University Press, Cambridge.
Heine, B. (2008) 'Contact-induced word order change without word order change', in P. Siemund and N. Kintana (eds.), Language Contact and Contact Languages, John Benjamins, Amsterdam/Philadelphia, 33-60.
Jaeger, T.F. (2006), 'Redundancy and syntactic reduction in spontaneous speech'. Unpublished doctoral dissertation, Stanford University, Stanford, CA.
Jaeger, T.F. and Tily, H. (2011) 'On language 'utility': Processing complexity and communicative efficiency', WIREs: Cognitive Science 2(3):323-335.
Jurafsky, D. (1996) 'A probabilistic model of lexical and syntactic access and disambiguation', Cognitive Science 20:137-194.
Keenan, E.L. (1975) 'Variation in Universal Grammar', in R. Fasold and R. Shuy (eds.), Analyzing Variation in English, Georgetown University Press, Washington, D.C., 136-148.
Keenan, E.L. and Hawkins, S. (1987) 'The psychological validity of the Accessibility

Hierarchy', in E.L. Keenan, Universal Grammar: 15 Essays, Croom Helm, London, 60-85.
Keenan, E.L. and Comrie, B. (1977) 'Noun phrase accessibility and Universal Grammar', Linguistic Inquiry 8: 63-99.
Kempson, R.M., Meyer-Viol, W., and Gabbay, D. (2001) Dynamic Syntax. Blackwell, Oxford.
King, R. (1969) Generative Grammar and Historical Linguistics. Prentice Hall, Englewood Cliffs, NJ.
Kirby, S. (1999) Function, Selection and Innateness. Oxford University Press, Oxford.
Kiss, K.E. (2002) The Syntax of Hungarian, Cambridge University Press, Cambridge.
Kuno, S. (1973) The Structure of the Japanese Language. MIT Press, Cambridge, Mass...
Kvam, S. (1983) Linksverschachtelung im Deutschen und Norwegischen. Max Niemeyer Verlag, Tuebingen.
Kwon, N., Gordon, P.C., Lee, Y., Kluender, R. and Polinsky, M. (2010) 'Cognitive and linguistic factors affecting subject/object asymmetry: An eye-tracking study of prenominal relative clauses in Korean', Language 86: 546-82.
Labov, W. (1972) Sociolinguistic Patterns. University of Pennsylvania Press, Philadelphia.
Labov, W. (1994) Principles of Linguistic Change 1: Internal Factors. Blackwell, Oxford.
Labov, W. (2007) 'Transmission and diffusion', Language 81: 344-387.
Levinson, S.C. (2000), Presumptive Meanings: The Theory of Generalized Conversational Implicature. MIT Press, Cambridge, Mass..
Levy, R. (2008) 'Expectation-based syntactic comprehension', Cognition 106: 1126-1177.
Lewis, R. and Vasishth, S. (2005) 'An activation-based model of sentence processing as skilled memory retrieval', Cognitive Science 29: 375-419.
Lightfoot, D.W. (1991) How to Set Parameters. MIT Press, Cambridge, Mass...
Lupyan, G. and Dale, R. (2010) 'Language structure is partly determined by social structure', PLoS One 5(1): e8559, doi:10.13l1/journal.pone.0008559
MacWhinney, B. (1982) 'Basic syntactic processes', in S. Kuczaj (ed.), Language Acquisition: Syntax and Semantics, Lawrence Erlbaum, Mahwah, N.J..
Manning, C.D. (2003) 'Probabilistic syntax', in R. Bod, J. Hay and S. Jannedy (eds.), Probability Theory in Linguistics, MIT Press, Cambridge, Mass., 289-341.
Matthews, S. and Yip, V. (1994) Cantonese: A Comprehensive Grammar, Routledge, London.
McMahon, A.M.S. (1994) Understanding Language Change. Cambridge University Press, Cambridge.
Miyamoto, E.T. (2006), 'Understanding sentences in Japanese bit by bit', Cognitive Studies: Bulletin of the Japanese Cognitive Science Society 13: 247-260.
Mobbs, I. (2008) ''Functionalism', the design of the language faculty, and (disharmonic) typology', MS, King's College Cambridge.
Newmeyer, F.J. (2005), Possible and Probable Languages: A Generative Perspective on Linguistic Typology. Oxford University Press, Oxford.
O'Grady, W. (2005) Syntactic Carpentry: An Emergentist Approach to Syntax. Lawrence Erlbaum Associates, Mahwah, New Jersey.
O'Grady, W. (2008) 'The emergentist program', Lingua 118: 447-464.
Phillips, C. (1996) 'Order and Structure', Ph.D. dissertation, MIT.

Primus, B. (1999) Cases and Thematic Roles: Ergative, Accusative and Active. Max Niemeyer Verlag, Tuebingen.

Prince, A. and Smolensky, P. (1993) Optimality Theory: Constraint Interaction in Generative Grammar. MIT Press, Cambridge, Mass..

Ross, J.R. (1967), 'Constraints on Variables in Syntax', Ph.D. dissertation, MIT.

Ross, M.D. (1996) 'Contact-induced change and the comparative method: Cases from Papua New Guinea', in M. Durie and M.D. Ross (eds.), The Comparative Method Reviewed: Regularity and Irregularity in Language Change, Oxford University Press, Oxford, 180-217.

Ross, M.D. (2001) 'Contact-induced change in Oceanic Languages in North-West Melanesia', in A.Y. Aikhenvald and R.M. Dixon (eds.), Areal Diffusion and Genetic Inheritance: Problems in Comparative Linguistics, Oxford University Press, Oxford, 134-166.

Saah, K.K. and Goodluck, H. (1995) 'Island effects in parsing and grammar: Evidence from Akan', Linguistic Review 12: 381-409.

Sadock, J. and Zwicky, A. (1985) 'Speech act distinctions in syntax', in T. Shopen (ed.), Language Typology and Syntactic Description, Vol.1: Clause Structure, Cambridge University Press, Cambridge.

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.

Shibatani, M. (1976) The Grammar of Causative Constructions. Academic Press, New York.

Sperber, D. and Wilson, D. (1995) Relevance: Communication and Cognition, 2nd edn.. Blackwell, Oxford.

Thomason, S. (2001) Language Contact – An Introduction. Edinburgh University Press, Edinburgh.

Thomason, S. and Kaufman, T. (1988) Language Contact, Creolization and Genetic Linguistics. University of California Press, Berkeley.

Tomlin, R.S. (1986) Basic Word Order: Functional Principles. Croom Helm, London.

Trudgill, P. (2011) Sociolinguistic Typology. Oxford University Press, Oxford.

van Oirsouw, R.A. (1987) The Syntax of Coordination. Croom Helm, London.

Wasow, T., Jaeger, T.F. and Orr, D.M. (2011) 'Lexical variation in relativizer frequency', in H. Simon and H. Wiese (eds.), Expecting the Unexpected, Exceptions in Grammar, de Gruyter, Berlin, 175-195.

Yamashita, H. (2002) 'Scrambled sentences in Japanese: Linguistic properties and motivation for production', Text 22: 597-633.

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

Yamashita, H. and Chang, F. (2006) 'Sentence production in Japanese' in M. Nakayama, R. Mazuka, and Y. Shirai (eds.), Handbook of East Asian Psycholinguistics, Volume 2, Japanese, Cambridge University Press, Cambridge.

Zipf, G. (1949) Human Behavior and the Principle of Least Effort. Hafner, New York.