ELSEVIER

New Ideas in Psychology 25 (2007) 87–107

NEW IDEAS IN PSYCHOLOGY

www.elsevier.com/locate/newideapsych

Processing typology and why psychologists need to know about it

John A. Hawkins

Cambridge University, UK

Available online 16 April 2007

Abstract

This paper illustrates an interdisciplinary research program based on cross-linguistic comparison that is of relevance for psychologists working on language processing, so-called "processing typology" [Hawkins, J. A. (1994). *A performance theory of order and constituency*. Cambridge: Cambridge University Press; (2004). *Efficiency and complexity in grammars*. Oxford: Oxford University Press]. Its most original feature is the hypothesis that patterns and preferences found in performance in languages with several structures of a given type (e.g. preferences among alternative word orders) are the same patterns and preferences one finds across languages in the fixed conventions of grammars that permit less variation (i.e. in fixed word orders). Data supporting this "performance–grammar correspondence hypothesis" are summarized. One of its consequences is that principles of performance can be used to make predictions for patterns of grammatical variation, while preferences in grammars become relevant for the testing of psycholinguistic ideas. Two proposed principles of ordering in performance, in terms of "end weight" and "memory cost", are criticized on the basis of cross-linguistic data. Both predict an asymmetry in ordering, whereby some category A precedes B. But end weight is not a valid cross-linguistic asymmetry, and memory cost cannot explain certain asymmetries for which it has been invoked when different language types are considered. The paper argues for greater mutual awareness between processing theorists and language typologists, for more consideration of non-European grammars and language types in psycholinguistics, and for a greater appeal to processing in the explanation of typological variation.
© 2007 Elsevier Ltd. All rights reserved.

1. Introduction

Processing typology is an interdisciplinary research program that examines grammatical patterns across languages in terms of language processing (Hawkins, 1994, 2004). It compares

E-mail address: jah91@cam.ac.uk.

0732-118X/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.newideapsych.2007.02.003

grammars with the patterns and preferences one finds in performance in languages with structural choices (between e.g. competing word orders or alternative relative clause structures) and it tests the hypothesis that the same principles underlie both sets of patterns. The potential interest of this approach for psychologists and for psychological models of performance is that grammars are viewed as conventionalizations of performance preferences. Grammars accordingly provide a new set of facts that psychologists can consult, in addition to their experimental and corpus data from the more familiar (generally European) languages, when setting up their production and comprehension models.

To give an initial flavor of this research program, consider relative clauses. These clauses may exhibit a 'gap' or a 'resumptive pronoun' strategy, in Hebrew and many other languages, giving alternations corresponding to *the students [that I teach 0]*) and *the students [that I teach them]*. A relative clause can also occur with and without a relative pronoun, as in English *the students [whom I teach]* vs. *the students [0 I teach]*. One of these strategies may be 'fixed' or 'conventionalized' in some environments, while there can be optionality and variation in others.

The selection from the variants in performance exhibits patterns. The retention of the relative pronoun in English is correlated with the degree of separation of the relative clause from its head, as Quirk (1957) first observed in his corpus of spoken British English. Zero relatives (i.e. those without relative pronouns) are preferred when adjacent to the head noun; and the greater the separation between the relative clause and its head, the greater the preference for the explicit relative pronoun (Hawkins, 2004). The Hebrew gap is similarly favored with smaller distances between the head noun filler and its gap than between the head noun and a resumptive pronoun (Ariel, 1999).

The distribution of gaps to resumptive pronouns in the fixed conventions of grammars also reveals patterns. The distribution follows the Keenan and Comrie (1977) Accessibility Hierarchy (AH) (Subjects > Direct Objects > Indirect Objects/Obliques > Genitives, cf. Comrie, 1989), with gaps preferred in the higher positions and explicit pronouns in the lower ones. More precisely, if a gap occurs low on the hierarchy, it occurs all the way up, and if a pronoun occurs high, it occurs all the way down (cf. Hawkins, 1999, 2004). Keenan–Comrie argued that this grammatical pattern was ultimately explainable by declining ease of processing down the AH. Examples of relative clauses formed on each of the positions of the AH are given in (1):

(1)	a	the professor*i* [that O*i* wrote the letter]	*SU*
	b	the professor*i* [that the student knows O*i*]	*DO*
	c	the professor*i* [that the student showed the book to O*i*]	*IO/OBL*
	d	the professor*i* [that the student knows his*i* son]	*GEN*

This line of reasoning has been generalized beyond relative clauses within the processing typology program. The preferred word orders in languages and structures with considerable freedom appear to be those that are grammaticalized in languages such as English with more fixed and basic orderings, as we will see in this paper. The basic hypothesis, the 'Performance-grammar correspondence hypothesis' (Hawkins, 2004), is defined in (2):

(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.

The PGCH defines a very different relationship between performance and grammars than the classic one presented in Chomsky (1965) and in most of his subsequent publications (e.g., Chomsky, 1986). The (competence) grammar is an important component of an overall performance model for Chomsky, but he has argued repeatedly that grammars are ultimately autonomous and independent of performance factors and that they are determined by an innate U(niversal) G(rammar). Processing typology is built on the opposite assumption of (2). It is an attempt to give a better account of cross-linguistic variation, to explain numerous exceptions to proposed universals, and to avoid unsupported appeals to innate grammatical parameters. The need for a different approach to cross-linguistic variation has been shown recently in the detailed critique of Chomskyan 'principles and parameters' by Newmeyer (2005). Newmeyer argues that these principles are not adequate empirically, when a wide range of languages is considered, and he advocates replacing them with principles of performance and processing along the lines of the PGCH.

The PGCH also provides a better fit with the findings of many branches of the language sciences, in which there is a growing awareness of this basic correspondence. For example, Haspelmath (1999) has 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 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 are provided for many of the basic constraints of that theory, some of them of an explicitly processing nature. A further development, Stochastic Optimality Theory (cf. Bresnan, Dingare, & Manning, 2001; Manning, 2003), is an explicit attempt to generate the preferences of performance ('soft constraints') as well as the grammatical conventions ('hard constraints') using the formal machinery of Optimality Theory, appropriately extended.

The PGCH is directly relevant for psychological models of performance. If the conventions of many grammars fly in the face of performance preferences predicted by a given model, then on the assumption that grammatical conventions have emerged out of performance preferences, they can be considered prima facie evidence against the model in question. In what follows we will see some examples of this sort.

I first give a brief summary of some major performance-grammar correspondences in the literature in order to convince psychologists reared on a diet of mainstream linguistics that this idea is not unreasonable, and indeed that it enjoys substantial support (Section 2). I then illustrate the processing typology research program by applying it to a set of grammatical and performance data that are very relevant for psycholinguistic models of production and comprehension (Sections 3–5). These data involve the distinction between 'symmetries' in ordering universals (Section 4) and 'asymmetries' (Section 5). A failure to recognize this distinction, and a lack of awareness of some of the basic linguistic generalizations that characterize each, has resulted in psychological models for which some basic facts of grammars provide counterevidence.

2. Some Examples of Observed Performance-Grammar Correspondences

The Keenan and Comrie (1977) AH illustrated in (1) has been much discussed in this context and was briefly summarized above. Grammatical cut-off points in relativization across languages follow the hierarchy, as does the distribution of gaps to resumptive pronouns, and Keenan–Comrie proposed their explanation in terms of declining ease of processing down each lower position of the hierarchy. As evidence they pointed to usage data from languages with many relativizable positions, especially English. The hierarchy correlated with declining corpus frequencies for relativizations down the hierarchy and with evidence of increasing processing load and working memory demands under experimental conditions (Keenan, 1975; Keenan & Hawkins, 1987). There have since been a number of psycholinguistic studies, in processing and acquisition and in a number of languages, that confirm the hierarchy as a complexity ranking, many of which are summarized in Hawkins (1999, 2004). Where minor departures from the hierarchy's predictions have been noticed, for example in Mandarin Chinese (Hsiao & Gibson, 2003), it has been argued that language-particular properties of the relative clause structure in the language in question are responsible and that the complexity explanation still holds.

More generally, there are many filler-gap dependency hierarchies for relativization and Wh-movement across grammars that correlate with increasing complexity in the permitted gap environments, measured in terms of the increasing distance between filler and gap and/or the increasing formal and semantic content of the structural domain linking filler and gap (Hawkins, 1999, 2004). The cut-off points in the grammatical conventions across languages (i.e. the distribution between grammatical and ungrammatical strings in the relevant languages) reflect the declining processing ease of relativizing on the relevant position, as measured in languages with numerous options. A supposedly innate constraint like 'subjacency' (Chomsky, 1981) excludes a set of particularly hard-to-process structures in this account. But there are languages that permit such hard-to-process structures (i.e. subjacency-violating ones like Akan, cf. Saah & Goodluck, 1995), and one can predict here that if such structures are grammatical, then all simpler filler–gap counterparts will be grammatical as well. See Hawkins (1999, 2004) for detailed testing of this idea and Hawkins (1994) for discussion of the conventionalization of grammatical rules in this and other areas in response to efficiency and complexity considerations in processing.

Reverse hierarchies across grammars for gaps in simpler environments and resumptive pronouns in more complex relativizations match the performance distribution of gaps to pronouns in languages like 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).

Parallel function effects (whereby the head of the relative matches the position relativized on in e.g. its subject or direct object relation) 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 the grammars of languages such as Basque and Hebrew (Aldai, 2003; Cole, 1976; Hawkins, 2004).

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 in accordance with their complexity (Hawkins, 1994, pp. 315–321)

(Nominative) subject before (accusative) object ordering is massively preferred in the performance of languages in which both S[ubject] before O[bject] 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; Tomlin, 1986; Primus, 1999).

Markedness hierarchies of case (Nom > Acc > Dat > Other) and number (Sing > Plur > Dual > Trial), etc., correspond to frequency of usage hierarchies in the performance of languages with rich morphological inventories (Croft, 1990; Greenberg, 1966; Hawkins, 2004).

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.

The distinction between zero agreement in local NP environments versus explicit agreement non-locally in the grammar of Warlpiri (where agreement is expressed through copying of the case marker on NP constituents) matches the environments in which zero and explicit agreement forms are preferred in performance in languages with choices. For example, the performance pattern of presence versus absence of relative pronouns co-indexed with the head in English (with the explicit relativizer favored under increasing distances from the head) matches the grammaticalized pattern of agreement in Warlpiri (with distant NP constituents requiring explicit agreement, while adjacent constituents have zero), cf. Hawkins (2001, 2004).

I believe that these are just the tip of a large iceberg of performance-motivated cross-linguistic patterns in grammars. And if these performance-grammar correspondences are valid, then any explanation for these phenomena that accounts only for grammars (such as the innateness of grammar) will be missing a significant generalization. Conversely psycholinguistic models of performance can be counterexemplified by grammatical as well as performance data, in the event that the conventions of grammars go against the predictions of these models. In the next section I turn to an important grammatical distinction between symmetries and asymmetries in word order universals, which will provide grammatical data of relevance to psycholinguistic models, while at the same time illustrating the usefulness of processing in understanding why grammars are the way they are.

3. Symmetries and Asymmetries in Word Order Universals

Across languages certain common categories {A, B} reveal symmetrical orderings: there are many languages with A + B and many languages with B + A. Other pairs of categories are consistently or preferably ordered asymmetrically A + B, regardless of the language type:

(3) Symmetries: A + B and B + A are productive.
Asymmetries: only A + B occurs or is significantly preferred.

For example, both Verb + Object (VO) and Object + Verb (OV) ordering are common cross-linguistically. The former is found in English, the latter in Japanese:

(4)	Taroo	wrote a letter.		(English)	S VO
(5)	Taroo ga	tegami o	kaita.	(Japanese)	S OV
	Taroo NOM	letter ACC	wrote		

Their distribution in different language samples is summarized in (6). Each is productive (for a discussion of the units counted in the different samples, languages (lgs) versus families (fams) and genera, cf. Dryer, 1989):

(6) VO and OV basic orders across languages (VO = SVO,VSO,VOS; OV = SOV,OVS)

VO/OV ratio:		
55%/45%	(78/64)	Greenberg's (1963) Appendix II (lgs & families)
48%/52%	(162/174)	Hawkins' (1983) Expanded Sample (lgs & fams)
54%/46%	(217/185)	Tomlin (1986) (lgs)
42%/58%	(79/111)	Dryer (1989) (genera)
39%/61%	(50/78)	Nichols (1992) (lgs)

An example of an asymmetry is WH-movement. Displaced WH-elements across languages almost always move to the beginning of their clauses, as in (7) from English, not to the end (Hawkins, 2002, 2004; Polinsky, 2002)

(7) a. Who did you say came to the party?
b. *You said came to the party who?

As a further example consider the fact that the topic in 'topic-prominent' languages like Japanese and Chinese generally stands to the left of its predication, not to its right:

(8) John wa gakusei desu (Japanese),
Speaking of John he is a student (Gundel, 1988; Kuno, 1973; Hawkins, 2004).

3.1. Two proposed asymmetries of ordering in psycholinguistic models

Discussions of asymmetric ordering have figured prominently in recent psycholinguistic models with respect to at least two phenomena. The first involves 'end weight': the preference for heavy and complex phrases (clauses, NPs, etc.) to be positioned at the end of their immediately containing clause or phrase, following lighter phrases (de Smedt, 1994; Stallings, 1998; Wasow, 1997, 2002). In English, for example, a heavy sentential subject as in (9a) and a heavy direct object NP as in (10a) are preferably postposed to the right, giving (9b) and (10b), respectively. This preference can be seen in both corpora and processing experiments (Erdmann, 1988, Stallings, 1998):

(9) a. s[That John got married yesterday] surprised Mary.
b. It surprised Mary s[that John got married yesterday]. (Heavy Last)

(10) a. I gave np[the antique that was extremely valuable and expensive] [to Mary]
b. I gave [to Mary] np[the antique that was extremely valuable and expensive]

One prominent explanation for this proposed asymmetry has been provided by de Smedt (1994). It builds on the kind of language production model formulated by Levelt (1989) and captures an intuition that is widely held in psycholinguistic thinking (cf. also Wasow,

1997, 2002). De Smedt argues that the on-line ordering of phrases reflects the processing time required by the Formulator to actually produce them in the parallel processes that are required for each. Short phrases can be formulated with greater speed and hence they are produced first in the competition between phrases and are ordered before heavy phrases. These 'weight effects' can be seen in a large number of ordering patterns in English and similar languages (Gibson, 1998; Hawkins, 1994, 2000, 2004), cf. Section 4.1.

A second set of proposed asymmetries that has received psycholinguistic attention involves the relative positioning of arguments of the verb, especially subject and direct object ordering. It has been observed that subjects are much preferred before objects in languages like German and Finnish in which both SO and OS occur as in the following examples from German:

(11) a. Die kleine blonde Frau küsste den roten Bären. (German)
the small blond lady kissed the red bear
b. Den roten Bären küsste die kleine blonde Frau.

Again there is a prominent psycholinguistic explanation for this preference, this time from Gibson (1998), building on the notion of 'memory cost' in on-line processing (a notion that goes back at least to Yngve, 1960). According to Gibson (1998, p. 59) object-first orders such as (11b) are more complex at the initial noun phrases because it is necessary to retain the prediction of a subject at this location. A direct object always co-occurs with a subject, and a clearly and unambiguously marked initial accusative object as in (11b) will activate the prediction that a (nominative) subject follows. Subject-first noun phrases do not activate the prediction of an object (clauses may be transitive or intransitive). Hence, according to Gibson, they require less working memory in on-line production and comprehension, and they can be expected to be more frequent and easier to process.

3.2. *End weight is not a valid cross-linguistic generalization*

If end weight is a consequence of the architecture of the production model, as de Smedt (1994) argues, one would expect this preference to be universal across languages, assuming that the language faculty (including the mechanisms of production) is universal. But this preference is far from universal. It is characteristic of languages like English that are 'head-initial': verbs occur initially in the verb phrase (VP) and early in the clause, nouns are initial or early in the NP, prepositions are initial in the PP, complementizers are initial in subordinate clauses, and so on, cf. Hawkins (1993, 1994). In other words, it is characteristic of languages that have VO, cf. (4) above. But in languages like Japanese, heavy phrases (clauses, NPs, etc) are preferably preposed to the left, as in (12b), and this can be seen both in corpus data and in on-line experiments (Hawkins, 1994; Yamashita & Chang, 2001):

(12) a. Mary ga [[kinoo John ga kekkonsi-ta to]s it-ta]vp (Japanese)
Mary NOM yesterday John NOM married that said, i.e.
Mary said that John got married yesterday.
b. [kinoo John ga kekkonsi-ta to]s Mary ga [it-ta]vp (Heavy First)

What this means is that weight effects are symmetrical, heavy-last or heavy-first depending on the language type, and this provides counterevidence to the speed of

formulation explanation proposed by de Smedt (1994). If short phrases systematically win the competition for early production in English, they should do so in Japanese as well. To the extent that they do not, then either the proposed architecture must be abandoned, or some competing mechanism needs to be proposed explaining why this outcome of the proposed universal processing architecture is no longer an outcome in these languages. The preference for longer phrases first is actually quite systematic in Japanese and other strongly 'head-final' languages, cf. Hawkins (1994), Yamashita and Chang (2001) and Section 4.2.

There is another complication posed by these weight effects for psycholinguistic models, especially for those that appeal to comprehension rather than production. In addition to their symmetry, weight effects are gradient. The preference for heavy-last or heavy-first depends on the *degree of difference* in weight between the heavy phrase and its lighter sisters, not on some absolute size that might exceed the capacity constraints of models such as Frazier's (1979, 1985), Just and Carpenter's (1992) or Gibson's (1998). The bigger the weight difference, the greater is the proportion of structures with heavy-last or heavy-first. For example, Heavy NP Shift in English (10b) applies in proportion to the weight difference between the NP and the PP, cf. Section 4.1 below.

This complicates the explanation for weight effects in terms of capacity constraints in working memory load since weight effects are visible among phrases with quite small weight differences (Hawkins, 1990, 1994). The observed patterns of preference in these cases fall within the proposed capacity constraints, and are unexplained by them, i.e. the constraints are too weak to account for weight effects in the data. They are also too strong since there are some language types with structures that will regularly exceed the capacity constraints (Hawkins, 1994).

3.3. Memory cost does not explain asymmetries of ordering

Gibson's (1998) subject before object preference fares better than end weight empirically. SO is generally preferred over OS in performance in different language types. E.g. only 4–5% of Japanese sentences in the corpus of Hawkins (1994, p. 145) had OSV as opposed to SOV. Basic orders across grammars also support it, as shown in Tomlin's (1986) language sample:

(13) Relative frequencies of basic word orders in Tomlin's (1986) sample (402 lgs)

SOV (168)	VSO (37)	VOS (12)	OVS (5)
SVO (180)			OSV (0)
87%	9%	3%	1%

Only 4% of these grammars have O before S. But when we look more closely at the morphological and syntactic properties of some of these languages, we see that an explanation in terms of the added memory cost of a predicting O before a predicted S cannot be correct.

First, a significant number of SOV and VSO languages have ergatively case-marked transitive subjects (Comrie, 1978; Primus, 1999). In these systems the (ergative) subject receives special and distinctive marking signaling a transitive agent acting on a patient. This marking predicts the co-occurrence of a required object. The (absolutive) object morphology, by contrast, does not predict a second (subject) argument since the absolutive

case is also found on intransitive subjects. An initial absolutive NP could be followed by an intransitive verb, therefore, or by an ergatively marked subject and a transitive verb. An initial absolutive in these languages is like an initial nominative in nominative-accusative systems: it does not predict a second following case-marked NP. The important point about these ergative-absolutive systems is that, in the great majority, the (predicting) ergative subject precedes the object in the most frequent and often basic order, despite the added memory cost of positioning the subject first. The North East Caucasian language Avar is typical. Example (14) is a transitive clause with the most common ordering of S and O, and (15) an intransitive. The absolutive is zero-marked in this and many other such languages:

(14)	Vas-as:	jas	j-ec:ula.	Avar (Anderson, 1976)
	boy-Erg	girl-Abs	Sg.Fem.Abs-praise, i.e. The boy praises the girl.	
(15)	Jas	j-ekerula.		
	girl-Abs	Sg.Fem.	Abs-run, i.e. The girl runs.	

A second problem for the memory cost explanation is that it is not the case that XY is always preferred over YX across languages when Y predicts X and not vice versa. Sometimes the non-predicting X is initial (English SO). Sometimes the predicting Y is initial, as in (14). Topic-marked phrases (e.g. Japanese *wa*) predict a predication and are also generally initial (8). Displaced WH-words predict a gap and are almost invariably initial (7). In all these cases the structures with greater memory cost are systematically preferred over those with less. Sometimes neither X nor Y predicts the other in these asymmetries. There is simply no consistent correlation between ordering asymmetries across languages and memory cost (cf. Hawkins, 2002, 2004 for further details).

There are further complications for any prediction-based explanation for ordering. One problem is that on-line predictions can both help processing, by activating co-occurrences at least one of which will be realized, and hinder it by adding to working memory. It is not clear, therefore, what ordering one should expect when an on-line prediction is made by one category for the co-occurrence of another. Empirically the predicting category is sometimes first (Avar (14) and Japanese topics (8), etc.) and sometimes last (accusatives following nominatives). The benefits and costs of an on-line prediction may simply be canceling each other out in these cases, making it an irrelevant consideration for ordering. I will suggest in Section 5 that there is in any case a different processing generalization that underlies these ordering asymmetries, involving the efficiency of the on-line assignments of properties to structures.

What we see in these brief examples is that psycholinguistic models have been set up on the basis of familiar European languages, with insufficient attention to cross-linguistic variation. There is a welcome trend towards experimental work on different language types (cf. e.g. Hsiao & Gibson, 2003; Matthews & Yeung, 2001; Saah & Goodluck, 1995; Yamashita & Chang, 2001, to cite just a handful of studies). But there needs to be much more awareness of cross-linguistic variation in psycholinguistics at this point. And since the study of diversity in grammars is decades ahead of performance studies of the relevant structures in more exotic languages, it is useful to examine grammars from a processing point of view at the same time that we broaden our data collection from language performance to include different language types.

4. Symmetries and domain minimization

Why are weight effects symmetrical and gradient? Consider a classic symmetry. Greenberg (1963) examined alternative verb positions across languages and their correlations with prepositions and postpositions in phrases corresponding to (16):

(16) 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 (----------)

(16a) is the English order, (16b) 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 (16c) and (16d). (17) summarizes the distribution in the database of Dryer's (1992) paper on the 'Greenbergian correlations' (cf. Hawkins, 2004, p. 124):

(17) 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 (17a) + (b) = 365/389 (94%)

From a parsing perspective, the adjacency of V and P guarantees the smallest possible string of words (indicated by the underlinings) for the recognition and construction of VP and of its two immediate constituents (ICs), i.e. V and PP. Non-adjacent V and P in (16cd) require longer and less efficient strings for the parsing of phrase structure. This leads to a hypothesis (which I first formulated in Hawkins, 1990): the recognition and construction of phrases and their combinatorial relations prefers the smallest possible string of words in processing (the principle of Early Immediate Constituents in Hawkins, 1990, 1994 and (21) below). More generally I have argued that the processing of all syntactic and semantic relations prefers minimal domains (Hawkins, 2004). This principle is defined in (18) and some sample predictions for performance and grammars will be illustrated in this section.

(18) *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 and maximal projections (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.

4.1. Syntactic MiD effects in the performance of head-initial languages

Recognizing how words and phrases combine together can typically be accomplished on the basis of less than all the words dominated by each phrase, cf. the preferred (16a) and (b) above. 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 now the following alternative orderings in post-verbal position in English, involving two PPs:

(19) 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, PP2 can be recognized on the basis of five words in (19a), compared with nine in (19b), assuming as I did for (16) that (head) categories like P project to mother nodes such as PP and enable the parser to immediately construct them and recognize them on-line. For comparable benefits within a Production Model, cf. Hawkins (2004, p. 106).

MiD 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. The principle of Early Immediate Constituents (a particular instance of MiD) is defined in (21):

(20) *Phrasal Combination Domain* (PCD): The PCD for a mother node M and its I(mmediate) C(onstituent)s consists of the smallest string of terminal elements (plus all M-dominated non-terminals over the terminals) on the basis of which the processor can construct M and its ICs.

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

In concrete terms EIC amounts to a preference for short before long phrases in head-initial structures like those of English, e.g. for short before long PPs in (19). 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 (19a) is $\frac{3}{5}$ or 60% (5 words required for the recognition of 3 ICs). The comparable ratio for (19b) is $\frac{3}{9}$ or 33% (9 words required for the same 3 ICs).

Structures like (19) 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):

(22)	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 kind of weight preference in English, e.g. Heavy NP Shift, cf. Hawkins (1994, p. 183), Wasow (1997, 2002), Stallings (1998). Wasow and Stallings provide experimental as well as corpus data supporting MiD. This principle is also very similar to Gibson's (1998) 'locality' principle for syntactic and semantic dependencies, and the experimental evidence that he cites for locality carries over to support domain minimization here.

A possible explanation for the distribution in (22) can be given in terms of reduced simultaneous processing demands in working memory. If, in (19a), the same phrase structure information can be derived from a 5-word viewing window rather than 9 words, then phrase structure processing can be accomplished sooner, there will be fewer additional (phonological, morphological, syntactic and semantic) decisions that need to be made simultaneously with this one, and less demands on working memory. (19a) is more efficient, therefore, by virtue of the fact that its phrasal combination domain is less complex than that of (19b). More generally we can hypothesize that the processing of all syntactic and semantic relations prefers minimal domains (Hawkins, 2004). There is no appeal here to capacity overload in working memory. Short before long is simply more efficient, for reasons of speed and less simultaneous processing effort—in English.

4.2. MiD effects in head-final languages

Why does Japanese do it differently? I suggest that this is because long before short orders provide minimal PCDs in head-final languages when constructing categories (V, P, Comp, case particles, etc.) are on the right. For example, if the direct object is a complement clause headed by the complementizer *to*, as in the Japanese example (12) repeated here, the distance between the complementizer and other constituents of the matrix clause in (12b), i.e. the subject *Mary ga* and the verb *it-ta*, is very short, just as short as it is in the mirror-image English translation *Mary said that … .* Hence the PCD for the matrix clause in (12b) is minimal. In (12a), by contrast, with the complement clause center-embedded, this matrix PCD proceeds all the way from *Mary ga* to *it-ta*, and is much longer.

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

A preference for (12b) 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, i.e. in alternations such as (23) (with *-o* standing for the accusative case particle, and PPm for a postpositional phrase with a head-final postposition):

(23) 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 the shorter and longer IC, respectively (i.e. with

weight the crucial distinction rather than phrasal type), these data are summarized in (24) (excluding the phrases with equal weights):

(24)	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)

The pattern here is the mirror image of that in (22): 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 been corroborated in the experimental and corpus data of Yamashita and Chang (2001), and it underscores the point made in Section 3.2: 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 (Hawkins, 1994, 2004).

4.3. Greenberg's word order correlations

Grammatical conventions across languages reveal the same degrees of preference for minimal domains, with the relative quantities of grammars reflecting the preferences, recall (17) above. An efficiency approach can account for this. It can also explain exceptions to majority patterns and to grammatical principles such as consistent head ordering (Hawkins, 1994, 2004).

The correlation pattern of (17), repeated here as (17′), has optimal IC-to-word ratios for (17′a) and (b) and much lower ratios for the inconsistently ordered (17′c) and (d) (Hawkins, 1990, 1994, 2004):

(17′) 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 (17′a) + (b) = 365/389 (94%) [data from Dryer's, 1992 sample]

(17′a) and (b) collectively account for 94% of languages in Dryer's sample. The adjacency of P and N within the PP is similarly preferred:

(25) a. pp[P np[N Possp]] = 134 (40%)
b. [[Possp N]np P]pp = 177 (53%)
c. pp[P [Possp N]np] = 14 (4%)
d. [np[N Possp] P]pp = 11 (3%)
EIC-preferred (25a) + (b) = 311/336 (93%) [data from Hawkins, 1983]

Other ordering universals point to the same preference for small and efficient PCDs, e.g. in noun-phrase-internal orderings corresponding to (26) in English:

(26) np[bright students s'[that Mary will teach]]
(26′) 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 now 12 logically possible orderings of Adj, N and S′ (ordered [C S] or [S C]). Just four of these have 100% IC-to-word ratios for the NP PCD (all 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), and these four account for the vast majority of languages. A small minority of languages are distributed among the remaining eight in proportion to their IC-to-word ratios measured on-line, cf. Hawkins (1990, 1994).

4.4. Explaining grammatical exceptions and unpredicted patterns

Dryer (1992) points out that there are systematic exceptions to Greenberg's correlations and to consistent head ordering when the category that modifies a head of phrase is a single-word item, e.g. an adjective modifying a noun (*yellow book*). Many otherwise head-initial languages have non-initial heads here (English is a case in point), 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 *books yellow with age* in English) there are good correlations with the predominant head ordering. Why should this be?

When a head category like V has a phrasal sister, i.e. a more complex and multi-word sister like 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. An intervening phrasal NP between V and P in (17′cd) makes the PCDs for mother VPs very long compared with their consistently ordered counterparts (17′ab), in which just 2 words suffice to recognize the 2 ICs. But when heads are separated by a non-branching single word, then the difference between, say, vp[V np[N Adj]] and vp[V np[Adj N]] 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. Similarly when there is just a one-word difference between competing domains in performance, e.g. in (22), 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. The NP-internal orderings of (26) across grammars are a case in point. Hierarchies of permitted center-embedded phrases provide further evidence. For example, in the environment pp[P np[__ N]] we have the following center-embedding hierarchy covering categories that can precede the noun (Hawkins, 1983):

(27)	Prep lgs:	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. As efficiencies decline the relative frequencies of prenominal orders in conventionalized grammatical rules also declines.

5. Asymmetries and Maximize On-line Processing

The point about the symmetries in Section 4 is that both orderings, A + B and B + A, can be efficient by the principle of MiD, and this provides a plausible explanation for the productivity of each. So what is it about the asymmetries (A + B/*B + A) that makes them asymmetrical?

I would like to point to a pattern that can be seen across grammars that suggests a second principle of efficiency, interacting with MiD, which I have called Maximize On-line Processing (Hawkins, 2002, 2004). One general property of many asymmetric orderings is that they involve an asymmetric dependency of B on A: properties to be assigned to B require access to A (cf. the definition of dependency given in (18) above), whereas those assignable to A do not require access to B, or require it less extensively and less frequently. In a clause with an antecedent and an anaphor, for example, *Johni washed himselfi*, the referent of *himself* depends on that assigned to the antecedent *John*. If *John* precedes *himself*, as it does in languages with SO, then the reference and coreferential index can be assigned to *himself* at the same time that the other (morphological and syntactic) properties of the anaphor are processed. But if *himself* precedes *John*, then reference and index assignment to the anaphor occur later than the other property assignments and their assignment must wait till the subject antecedent has been processed and be assigned retrospectively. Object before subject orders are less efficient in this respect. Similarly if the gap were to precede the WH-word in (7b) [*you said O*i *came to the party*] *who*i, there would be a delay in assigning the subject argument to *came*.

This is the essence of the Maximize On-line Processing idea. A large number of asymmetric orderings appear to involve such asymmetric dependencies of B on A, with the result that the prior positioning of B would delay property assignments on-line. This is defined in (28):

(28) *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.

Misassignments on-line are 'garden path' sentences, which have been consistently shown to cause processing difficulty (Frazier, 1985; MacDonald, Pearlmutter, & Seidenberg, 1994). Examples are: ***The horse raced past the barn*** *fell*; *I believe* ***the boy*** *knows the* answer; and *While Mary was* ***reading the book*** *fell down* (the misassignment areas are shown in bold). Misassignments are inefficient: the misassigned properties make no contribution to the ultimate syntactic and semantic representation of the sentence and their repair requires processing effort. Unassignments on-line are also inefficient. Unassignments arise when a form B cannot receive its full set of properties on-line because some property of B, P, is dependent for its assignment on form A which follows B (B + A), see Marcus (1980) on 'look ahead' in parsing. The reverse ordering (A + B) may permit a full assignment of properties to B as it is processed, as in the *John washed himself* example.

Unassignments and misassignments reduce the ratio of On-line Properties to Ultimate Properties assigned (the OP-to-UP ratio). A metric for quantifying this is proposed in Hawkins (2004, p. 55):

(29) *On-line Property to Ultimate Property Ratio*: The OP-to-UP ratio is calculated at each word X within a connected set of words {...X...} whose property assignments differ between two competing structures S and S′. The cumulative number of properties assignable at each X is divided by the total number of properties to be assigned in the connected set in the ultimate representation of each structure, and the result is

expressed as a percentage (e.g. 4/20 = 20%, 8/20 = 40%, 10/20 = 50% for successive words, etc). The higher these on-line percentages, the more efficient is structure S or S′, since more properties are assigned earlier.

5.1. Topic before dependent predication

The asymmetric positioning of topic phrases, in languages with conventionalized topic positions, suggests an explanation along the lines of MaOP (28). There are numerous dependencies holding between predication and topic in e.g. Chinese and Japanese, and these dependencies are either fully or predominantly asymmetrical. The topic has independent reference (cf. the definiteness and universal reference effects summarized in Kuno, 1973; Li & Thompson, 1981; Matthews & Yip, 1994; Tsao, 1978), but the predication typically depends on access to the topic for numerous aspects of its interpretation. If predication and topic ordering were reversed, there would be little impact on the on-line processing of the topic, but significant aspects of the interpretation of the predication would be delayed, resulting in unassignments and possible misassignments. These involve: the assignment of arguments to a verb in the predication; various enrichments of an argument in the predication; and various enrichments of the predicate within the predication.

The following examples are taken from Tsao's (1978) detailed study of topics in Mandarin Chinese (Tsao, 1978) and they illustrate each of these possibilities: **argument assignments** to a verb in the predication made by reference to the topic (the topic in this and subsequent examples is shown in bold):

(30) **Jang San (a)**, dzwo-tyan lai kan wo. (argument assignment)
Jang San (Topic Part), yesterday (he) came (to) see me.

argument enrichments whereby the topic provides a **possessor** (31), **class** (32), **set** (33) or **restrictive adjunct** (34) relative to which an argument in the predication is interpreted:

(31) **Jei-ge ren (a)**, tounau jyandan. (argument enrichment: possessor-possessed)
This-Classif man (Topic Part), (his) mind (is) simple.

(32) **Wu-ge pinggwo (a)**, lyang-ge hwai-le. (argument enrichment: class-member)
Rice-Classif apples (Topic Part), two-Classif (are) spoiled.

(33) **Ta-de san-ge haidz (a)**, yi-ge dang lyushr. (argument enrichment: set-member)
His three-Classif children (Topic Part), one-Classif serve-as lawyer.

(34) **Jei-jyan shr (a)**, wo-de jingyan tai dwo-le. (argument enrichment: restrictive adjunct)
This-Classif matter (Topic Part), my experience too many.

predicate enrichments whereby the topic provides a **location** (35), **time** (36), or **cause** (37) adjunct, or a domain for **superlative** (38) interpretation relative to which the predication is interpreted:

(35) **Nei kwai tyan (a)**, daudz jang de hen da. (predicate enrichment: location)
That piece land (Topic Part), rice grows Part very big (in it).

(36)	**Dzwo-tyan (a)**, Jang San lai kan wo. Yesterday (Topic Part), Jang San came (to) see me.	(predicate enrichment: time
(37)	**Weile jei-ge haidz**, wo bu jr chr-le dwoshau ku. For (/on account of) this-Classif child, I have endured much hardship.	(predicate enrichment: cause)
(38)	**Yu (a)**, wei-yu syandzai dzwei gwei. Fish (Topic Part), tuna is now the most expensive.	(predicate enrichment: superlative domain)

Reversing predication and topic would result in on-line unassignments and misassignments. In (31), for example, it would be unclear at the predication whose mind was intended; in (34) the absence of the restriction imposed by the topic would lead to an overly general interpretation on-line that could be untrue (my experience in general vs. my experience in this matter); in (38) the expensiveness of tuna must be interpreted relative to fish, not e.g. food in general, and this restricted interpretation cannot be assigned on-line when fish follows.

These asymmetries in dependency lead to a predicted topic before predication ordering, by MaOP (28). Across languages, argument enrichment and predicate enrichments structures (in which the dependencies between predication and topic are fully asymmetric) do indeed appear to be ordered entirely topic before predication (Gundel, 1988), i.e. for gap-containing non-dislocation predications. Argument assignment dependencies (which are predominantly but not fully asymmetric, since a topic can also be dependent on the predication for the assignment of its thematic role) are preferably topic before predication, with occasional reports of predication before topic languages, e.g. in Ojibwa (Tomlin & Rhodes, 1979).

5.2. *Subjects before objects*

Why are subjects preferred before objects? In Section 3.3 I critiqued the memory cost idea as an explanation for this asymmetry. An alternative can be given in terms of asymmetric dependencies and MaOP (28).

Objects are asymmetrically dependent on subjects for numerous property assignments, and if objects precede, these dependencies are unassignable on-line. Anaphoric binding has already been mentioned. It is the object that is systematically dependent on the subject in *Johni washed himselfi*, and if the object precedes (*Himselfi Johni washed*) there is delayed reference and index assignment. Objects are also referentially dependent on subjects in many other ways, e.g. in quantifier scope interpretations. Compare the sentences of (39) with those of (40):

(39)	a.	All the students read two books.
	b.	Many students read two books.
	c.	Three students read two books.
(40)	a.	Two books all the students read.
	b.	Two books many students read.
	c.	Two books three students read.

When *two books* follows the subject in (39), the preferred interpretation varies with the choice of subject (the so-called narrow scope interpretation); when it precedes in (40) the

preferred interpretation is *in*dependent of a subject that has not yet been processed (the wide scope interpretation), i.e. two specific books.

There has also been considerable discussion in linguistics of the 'thematic roles' that are assignable to subjects and objects, see Dowty's (1991) 'Proto-Agents and Proto-Patients and Primus' (1999) extensions of this work to make it applicable to languages with different types of case marking, verb agreement and configurationality properties. This work is relevant in this on-line processing context since it has been argued that the semantic role of the object is dependent on that of the subject. In *The hunter shot the deer* the transitive Patient (*the deer*) is acted on by the Agent (*the hunter*). More generally, the Patient is argued to be asymmetrically dependent on the Agent for its interpretation and this latter must be accessed for full interpretation of the Patient role. The Agent, by contrast, could be carrying out the activity described by the verb (*shooting*), whether he affected the particular Patient or not, but the latter's role as a Patient does depend on the Agent carrying out the action in question. Ordering the subject before the object makes the subject available for processing first, and dependent properties of the object can then be assigned immediately as they are processed, allowing for more efficient property assignments on-line.

The facts summarized in Section 3.3 and in this section suggest that it is not on-line predictions per se, and memory cost, that account for the asymmetrical ordering preference of subjects before objects. Rather it is the many (asymmetrical) dependencies of objects on subjects and the resulting on-line processing advantages of positioning the subject first. When the dependencies can be removed, as in the wide scope quantifier interpretations of (40), objects are preferred first and the normal linear ordering is reversed. This alternative explanation is supported by a wealth of other asymmetric dependency/asymmetric ordering correlations in performance and grammars (Hawkins, 2002, 2004).

6. Conclusions

I conclude that current psycholinguistic models are being set up with insufficient attention to cross-linguistic variation. End weight is not a cross-linguistically valid asymmetry (Section 3.2). Gibson's memory cost explanation for subject before object ordering lacks cross-linguistic generality (Section 3.3). On the other hand, attempts by linguists to explain and predict grammatical variation can benefit from the inclusion of processing ideas in the very core of their grammatical theorizing, for example when explaining the Greenbergian head ordering correlations (Section 4.3), and when explaining asymmetric orderings (Sections 5.1 and 5.2). A processing approach to grammatical universals can avoid the stipulations of current formal grammars, make predictions from performance data for hypothesized grammatical universals and suggest explanations for exceptions to purely grammatical principles (Section 4.4).

At the same time the conventionalized data of grammars can lead to testable predictions for performance. I first proposed the principle of Early Immediate Constituents (21), an instance of Minimize Domains (18), on the basis of the Greenbergian word order correlations in Hawkins (1990) (Section 4.3). The consequences predicted for performance were subsequently tested in Hawkins (1994) and have been tested further in publications cited in this paper. The asymmetric ordering idea presented in Section 5 was initially suggested by grammatical ordering conventions such as WH-movement (Hawkins, 2002, 2004) and topic before predication orders in Chinese and Japanese.

The moral I draw from this is that we need to look at different kinds of languages at this point, especially those outside of Europe, and we need to let broader and better empirical generalizations from both performance and grammars suggest improved processing principles. Two such principles have been proposed here: Minimize Domains (18); and Maximize On-line Processing (28). The data predicted by, and supporting these principles, suggest more generally that there is a profound correlation between performance and the conventionalized rules of grammar: patterns and preferences of performance in languages possessing several structures of a given type appear to be the same patterns and preferences found in the fixed conventions of grammars, in languages with fewer structures of the same type. This is the essence of the Performance-Grammar Correspondence Hypothesis (PGCH) formulated in (2). In the present context I have tried to illustrate some of its predictions in the area of word order. Many more phenomena are discussed in Hawkins (2004) along with some of the general issues that they raise for current linguistic and psycholinguistic theories.

References

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

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

Anderson, S. (1976). On the notion of subject in ergative languages. In C. N. Li (Ed.), *Subject and topic* (pp. 1–24). New York: Academic Press.

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

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. (1986). *Knowledge of language: Its nature, origin, and use*. New York: Praeger.

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. (1978). Ergativity. In W. Lehmann (Ed.), *Syntactic typology*. Austin: University of Texas Press.

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

Croft, W. (1990). *Typology and universals*. Cambridge: Cambridge University Press.

de Smedt, K. J. M. J. (1994). Parallelism in incremental sentence generation. In G. Adriaens, & U. Hahn (Eds.), *Parallelism in natural language processing*. Norwood, NJ: Ablex.

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

Dryer, M. S. (1989). Large linguistic areas and language sampling. *Studies in Language*, *13*, 257–292.

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

Erdmann, P. (1988). On the principle of "weight" in English. In C. Duncan-Rose, & T. Vennemann (Eds.), *On language, rhetorica, phonologica, syntactica: A festschrift for Robert P. Stockwell from his friends and colleagues* (pp. 325–339). London: Routledge.

Frazier, L. (1979). *On comprehending sentences: Syntactic parsing strategies*. Bloomington: Indiana University Linguistics Club.

Frazier, L. (1985). Syntactic complexity. In D. Dowty, L. Karttunen, & A. Zwicky (Eds.), *Natural language parsing*. Cambridge: Cambridge University Press.

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

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.
Gundel, J. K. (1988). Universals of topic-comment structure. In M. Hammond, E. Moravcsik, & J. Wirth (Eds.), *Studies in syntactic typology*. Amsterdam: John Benjamins.
Haspelmath, M. (1999). Optimality and diachronic adaptation. *Zeitschrift für Sprachwissenschaft*, *18*, 180–205.
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. (1994). *A performance theory of order and constituency*. Cambridge: Cambridge University Press.
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. (2001). Why are categories adjacent? *Journal of Linguistics*, *37*, 1–34.
Hawkins, J. A. (2002). Symmetries and asymmetries: Their grammar, typology and parsing. *Theoretical Linguistics*, *28*, 211–227.
Hawkins, J. A. (2004). *Efficiency and complexity in grammars*. Oxford: Oxford University Press.
Hsiao, F., & Gibson, E. (2003). Processing relative clauses in Chinese. *Cognition*, *90*, 3–27.
Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*(1), 122–149.
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., & 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 (Ed.), *Universal grammar: 15 essays*. London: Croom Helm.
Kirby, S. (1999). *Function, selection and innateness: The emergence of language universals*. Oxford: Oxford University Press.
Kuno, S. (1973). *The structure of the Japanese language*. Cambridge, MA: MIT Press.
Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
Li, C. N., & Thompson, S. A. (1981). *A functional reference grammar of Mandarin Chinese*. Berkeley: University of California Press.
MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*(4), 676–703.
MacWhinney, B. (1982). Basic syntactic processes. In S. Kuczaj (Ed.), *Language acquisition: Syntax and semantics*. Mahwah, NJ: Erlbaum.
Manning, C. D. (2003). Probabilistic syntax. In R. Bod, J. Hay, & S. Jannedy (Eds.), *Probability theory in linguistics* (pp. 289–341). Cambridge, MA: MIT Press.
Marcus, M. P. (1980). *A theory of syntactic recognition for natural language*. Cambridge, MA: MIT Press.
Matthews, S., & Yeung, L. Y. Y. (2001). Processing motivations for topicalization in Cantonese. In K. Horie, & S. Sato (Eds.), *Cognitive-functional linguistics in an East Asian context* (pp. 81–102). Tokyo: Kurosio Publishers.
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.
Matthews, S., & Yip, V. (1994). *Cantonese: A comprehensive grammar*. London: Routledge.
Newmeyer, F. J. (2005). *Possible and probable languages*. Oxford: Oxford University Press.
Nichols, J. (1992). *Linguistic diversity in space and time*. Chicago: University of Chicago Press.
Polinsky, M. (2002). Efficiency preferences: Refinements, rankings, and unresolved questions. *Theoretical Linguistics*, *28*(2), 177–202.
Primus, B. (1999). *Cases and thematic roles*. Tübingen: Max Niemeyer Verlag.
Quirk, R. (1957). Relative clauses in educated spoken English. *English Studies*, *38*, 97–109.
Saah, K. K., & Goodluck, H. (1995). Island effects in parsing and grammar. *Linguistic Review*, *12*, 381–409.
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, USC.
Tomlin, R. S. (1986). *Basic word order: Functional principles*. London: Routledge.

Tomlin, R. S., & Rhodes, R. (1979). An introduction to information distribution in Ojibwa. In *Papers from the 15th regional meeting of the Chicago linguistic society* (pp. 307–320).

Tsao, F.-F. (1978). *A functional study of topic in chinese*. Taipei: Student Book Company.

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

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

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

Yngve, V. H. (1960). A model and an hypothesis for language structure. *Proceedings of the American Philosophical Society*, *104*, 444–466.