
Efficiency and Complexity in Grammars: Three General Principles¹

JOHN A. HAWKINS

1 Introduction

An interesting general correlation appears to be emerging between variation data in performance, in languages that dispose of several structures of a given type, and the fixed conventions of grammars, in languages with fewer structures of the same type. The performance data come from corpus frequencies and processing experiments, the grammatical data from our expanding samples of languages and from the growing number of languages that have now been subjected to in-depth analysis.

The Keenan-Comrie Accessibility Hierarchy was an early example of such a correlation, whereby an implicational hierarchy in cross-linguistic relativization possibilities (SU>DO> IO/OBL>GEN) was argued to correlate with ease of relativization and with relativization preferences in a language (English) with many relativizable positions (Keenan and Comrie 1977, Keenan 1975, Keenan and Hawkins 1987). The preferred word orders in languages with choices have also been argued to be those that are productively grammaticalized in languages with fixed orders, and in almost exact proportion to their degree of preference, in Hawkins (1994, 1998a, 2001).

The goal of this paper is to define a hypothesis concerning the relationship between performance and grammars, and to propose three general prin-

¹ I would like to thank Jeff Elman for a very helpful and detailed review of the original version of this paper.

principles of efficiency and complexity that appear to underlie the preferences of performance and their correlations with grammars. The hypothesis will be referred to as the Performance-Grammar Correspondence Hypothesis (PGCH). The three principles are supported by data from several areas of syntax and morpho-syntax. If this hypothesized correspondence between performance and grammars is even partially correct, it needs to be seriously pursued in order to see how much of core syntax and of cross-linguistic syntactic variation can be derived from whatever explains the preferences of performance. If there is a common explanation for both, this will have far-reaching consequences for the ultimate explanatory basis of grammatical conventions.

The PGCH is defined in (1):

- (1) PERFORMANCE-GRAMMAR CORRESPONDENCE HYPOTHESIS
Grammars have conventionalized syntactic structures in proportion to their degree of preference in performance, as evidenced by frequency of use and ease of processing.

From (1) we can derive a number of predictions, including those of (2):

- (2) Predictions of the PGCH
- a. If a structure A is preferred over an A' of the same structural type in performance, then A will be more productively grammaticalized, in proportion to its degree of preference; if A and A' are more equally preferred, then both A and A' will be productive in grammars.
 - b. If there is a preference ranking $A > B > C > D$ among structures of a common type in performance, then there will be a corresponding hierarchy of grammatical conventions (with cut-off points and declining frequencies of languages).
 - c. If two preferences P and P' are in (partial) opposition, then there will be grammatical variation, with both P and P' being conventionalized, each in proportion to the degree of motivation for that preference in a given language type.

In order to test these predictions we need performance data from languages with a plurality of structures of a common type, e.g. different orderings of the same phrasal constituents, or relativizations on different positions of a clause using the same relative clause type (e.g. a Filler-Gap construction), and so on. We then need cross-linguistic grammatical data from languages

with fewer structures of these types, i.e. with more conventionalized and fixed phrasal orderings, and with relativizations on fewer positions. If the degree of preference in performance matches the relative productivity of structures in the cross-linguistic grammatical conventions, then the PGCH will be supported. If not, it will not be.

More generally, we need to identify general principles of preference in performance and to search for corresponding principles in grammars. In this way we can rise above a discussion of individual structures and languages, with their possible idiosyncracies, and focus on general properties of preference that hold for a large number of structures in a large number of languages. This paper will attempt to do this by outlining three general principles.

2 Principle 1: MINIMIZE DOMAINS

One clear principle of efficiency and complexity involves the size of the syntactic domain within which a given grammatical relation can be processed. How great is the distance separating interrelated items and how much material needs to be processed simultaneously with the processing of the relation? Hawkins (1994, 1998a, 1999a, 2001) argues that there is a correlation between performance and grammar with regard to domain sizes. In those languages and structures in which domain sizes can vary in performance, we see a clear preference for the smallest possible domains. In those languages and structures in which domain sizes have been grammatically fixed, we see the same preference in the conventions.

(3) MINIMIZE DOMAINS

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 will be 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 is defined here as: two categories A and B are in a relation of combination iff they occur within the same syntactic mother phrase or maximal projection. And DEPENDENCY is defined as: two categories A and B are in a relation of dependency iff the processing of B requires access to

also construct M). The greater efficiency of (4a) can then be captured by dividing the number of ICs by the number of words in each PCD and by expressing the result as a percentage, following the procedure in Hawkins (1994: ch.3).³ The higher the percentage, the more efficient the ordering is, since the same constituency information can be assigned on the basis of less terminal material:

(4a) VP PCD: IC-to-word ratio of $3/5 = 60\%$

(4b) VP PCD: IC-to-word ratio of $3/9 = 33\%$

My earlier principle of Early Immediate Constituents (EIC, cf. Hawkins 1990, 1994) is now subsumed under Minimize Domains and can be redefined as follows:

(6) EARLY IMMEDIATE CONSTITUENTS (EIC)

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.

EIC's predictions remain as they were.

A possible explanation for the EIC preference is that small PCDs reduce simultaneous processing in working memory. If, in (4a), the same VP constituency information can be assigned on the basis of five words, whereas nine words are required in (4b), then the production and comprehension of phrase structure 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 (cf. Hawkins 1999b, 2001).

2.1.1 Performance Data Testing EIC

For a language like English, EIC defines a quantitative preference for shorter PPs before longer ones overall, and the degree of this preference is predicted to be directly proportional to the length difference between them. Structures like (4a) and (4b) were examined in which the two PPs were

³ See Hawkins (1994: Ch.3) and also Wasow (2002) for discussion and testing of alternative ways of quantifying the EIC preference, in terms of IC-to-word ratios and IC-to-nonIC ratios using different non-IC nodes (e.g. phrasal nodes only or all non-terminal nodes). Wasow (ibid.) concludes that the different quantification procedures are statistically undistinguishable and that the simpler IC-to-word ratio fares as well as any.

freely permutable with truth-conditional equivalence (Hawkins 1999b).⁴ Only 15% were ordered long before short, with 67% short before long and 18% of equal length (n=71). The short before long orders plus the equal length PPs (in which either ordering is as efficient as it can possibly be) collectively account for 85% of the data. Moreover, within the 15% of long before short PPs, almost two thirds involved only a one-word difference in length and there was a gradual reduction in long before short orderings the bigger the length difference, as predicted by (6). This is shown in (7), in which the top row gives the figures for short before long, PP¹ before PP², and the second row long before short.

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

Data from Japanese reveal an equally principled set of preferences, but for the mirror-image long before short pattern. This is predicted by EIC. Postposing a heavier phrase to the right in English shortens PCDs and increases IC-to-word ratios. Preposing heavy phrases in a head-final language has the same effect, since the relevant constructing categories (V for VP, P for PP, etc) are now on the right. In a structure like $[[[NP\ P]_{PP1}, [NP\ P]_{PP2}] V]_{VP}$, the PCD for VP will proceed from the first P(ostposition) encountered to the verb, and will be shorter if the shorter PP₁ is adjacent to the verb than if the longer PP₂ is adjacent. The preferred pattern overall should be long before short in Japanese, therefore, and the degree of this preference should increase with increasing relative weights. In this way the time course from construction of the first PP to the VP-final verb will be faster than if the long PP is adjacent to V.

Consider some illustrative data collected by Kaoru Horie and involving orderings of $[[NP_o, PP_m] V]$, where NP_o stands for a direct object containing an accusative case particle *o*, in combination with a postpositional phrase (PP_m, i.e. with right-peripheral projection from P to its mother PP, hence the *m* subscript on the right). I assume that *o* is the projecting and constructing category for this case-marked NP, paralleling the final postposition in PP and the final V in VP, and that these heads have left-branching phrasal sisters that are processed bottom-up. I.e. the NP within PP_m is

⁴ The corpus that was used for sequences of PPs following intransitive verbs consisted of 500 pages of written English, comprising 200 pages of fiction, 225 pages of nonfiction, and 75 pages of a diary. Eight different texts were used, and the full references and page numbers are given in Appendix 1 of Hawkins (1999b).

processed before PP itself is constructed on the basis of P, the NP within NP_O before the case particle is processed projecting to a case-marked NP, and both are processed before the containing VP is constructed on the basis of V. The distance and time course from the first constructing category (P or *o*) to V is then shorter when the longer sister phrase precedes the shorter one. I.e. [PP_m NP_O V] is preferred when PP_m>NP_O, and [NP_O PP_m V] is preferred when NP_O>PP_m. An example of the relevant sentence type is given in (8):

- (8) Tanaka ga {[sono hon o] [Hanako kara]} katta.
 Tanaka SU that book OBJ Hanako from bought,
 ‘Tanaka bought that book from Hanako.’

The data (cf. Hawkins 1994:152) reveal that each additional word of relative weight resulted in a higher proportion of long before short orders, as shown in (9), in which IC₂ refers to the longer phrase, whether NP_O or PP_m, and IC₁ to the shorter.⁵

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

An additional 91 sequences had ICs of equal length (total n = 244). The preferred long before short sequences plus those with equal lengths (in which either order was as efficient as it could possibly be) amounted to 201/244 = 82%, the dispreferred short before long to 43/244 = 18%. A full 30 of these 43 involved a length difference of only 1-2 words and hence EIC defined its weakest possible dispreference in these cases. Further support for the long before short preference in Japanese is given in Yamashita and Chang (2001).

2.1.2 EIC in Grammars

Grammatical conventions across languages reveal the same EIC preference for minimal domains. Minimal PCDs are preferred in the vast majority of languages, and as PCD sizes increase, the relevant orders are conventionalized in progressively fewer languages and structures.

⁵ The Japanese data in (9) are taken from three written texts referenced in Hawkins (1994).

The cross-category ordering correlations in Greenberg's (1963) universals, two of which are presented in (10) and (11), illustrate the minimal domain preference in VP and PP respectively:

- (10) a. $VP[V_{PP}[P\ NP]]$ = 161 (41.4%)
 IC-to-word: 2/2 = 100%
 b. $[[NP\ P]_{PP}\ V]_{VP}$ = 204 (52.4%)
 IC-to-word: 2/2 = 100%
 c. $VP[V\ [NP\ P]_{PP}]$ = 18 (4.6%)
 IC-to-word: 2/4 = 50%
 d. $[_{PP}[P\ NP]\ V]_{VP}$ = 6 (1.5%)
 IC-to-word: 2/4 = 50%

Assume: V = 1 word; P = 1; NP = 2
 EIC-preferred (10a)+(10b) = 365/389 (94%)

- (11) a. $PP[P\ _{NP}[N\ POSSP]]$ = 134 (39.9%)
 b. $[[POSSP\ N]_{NP}\ P]_{PP}$ = 177 (52.7%)
 c. $PP[P\ [POSSP\ N]_{NP}]$ = 14 (4.2%)
 d. $[_{NP}[N\ POSSP]\ P]_{PP}$ = 11 (3.3%)

EIC-preferred (11a)+(11b) = 311/336 (93%)

The adjacency of V and P, and of P and N, guarantees the smallest possible domain for recognition and production of the two ICs in question (V and PP within VP, P and NP within PP): two words suffice, hence 100% IC-to-word ratios. In the non-adjacent domains of the (c) and (d) orders, the ratios are significantly lower and exemplifying languages are significantly fewer. The preferred (a) and (b) structures collectively account for 94% or 93% of all languages (see Dryer 1992 for the data of (10), Hawkins 1983 for (11)). These data support prediction (2a) above.

It is patterns like this that have motivated the head-initial (or VO) and head-final (OV) parameters in typological and generative research. The two language types are mirror images of one another, and EIC offers an explanation for this: both (a) and (b) are optimally efficient. They are equally good orderings for phrase structure recognition, and we expect both to be productive by prediction (2a) above.

When non-optimal domains are grammatically permitted they appear to be arranged in hierarchies whose internal organization reflects the complexity of the relevant domain, as predicted by (2b). For example, it has been well-known since Miller and Chomsky (1963), Bever (1970) and Kimball (1973) that center embeddings are hard structures to process. EIC predicts

that the larger a center-embedded constituent is relative to its sisters, the greater will be its degree of dispreference. For example, the heavier an NP relative to a PP in $_{VP}[V NP PP]$ and the longer the PCD for this containing VP, the more infrequent and the harder to process this structure is in English performance, as shown in Hawkins (1994), Stallings (1998) and Wasow (2002) using both frequency data and processing experiments. Correlating with this are hierarchies of tolerance for center embeddings of different degrees of complexity in cross-linguistic grammatical conventions, cf. Hawkins (1994).

For example, in the center-embedding environment between a preposition and the head noun within a sister NP, i.e. $_{PP}[P NP[_ N]]$, we find the following implicational pattern:

- (12) If a language is prepositional, then if RelN then PosspN [Exceptionless]; if PosspN then AdjN [Strongly supported in Hawkins (1983)]; and if AdjN then DemN [Massively supported but e.g. Sango is exceptional, cf. Dryer (1993)].

and the following distribution of preposed to postposed noun modifiers in the prepositional languages of Hawkins' (1983) sample:

| | | | | |
|----------------|-------|-----|-------|-----|
| (13) Prep lgs: | RelN | 1% | NRel | 99% |
| | PospN | 12% | NPosp | 88% |
| | AdjN | 32% | NAdj | 68% |
| | DemN | 49% | NDem | 51% |

Relative clauses are on aggregate longer than possessive phrases, which are on aggregate longer than single-word adjectives. As center embedded categories of increasing aggregate length intervene between P and N, the distance between them increases and IC-to-word ratios decline. As they do so the, the numbers of attested languages decline, and the proportion of languages with postposed rather than preposed modifiers increases. Indeed the only center-embedded category that is productive at all is the single-word adjective (at 32%). Prenominal possessive phrases are found in only 12% of prepositional languages, and prenominal relatives in only 1%. Demonstrative determiners construct the same phrase (NP) as head nouns in the theory of Hawkins (1993, 1994) and both $[P [DemN]]$ and $[P [NDem]]$ are therefore equally efficient.

Grammatical structures are avoided across languages in proportion to EIC's dispreference, therefore, resulting in this center-embedding hierarchy.

2.2 Domain Minimization and Adjacency

More generally, principle 1 (MINIMIZE DOMAINS) can be used to explain adjacency effects in syntax. Categories that prefer adjacency are argued in Hawkins (2001, in preparation) to be those that contract more combinatorial and/or dependency relations with one another, and whose processing domains are minimal when adjacent. Hawkins (*ibid.*) proposes the following hypothesis:

(14) ADJACENCY HYPOTHESIS

Given a phrase {H, {X, Y}}, H = a head category and X and Y phrases that are potentially adjacent to H, then the more combinatorial and dependency relations whose processing domains can be minimized when X is adjacent to H, and the greater the minimization difference between adjacent X and adjacent Y in each domain, the more H and X will be adjacent.

Consider again the PP sequences in the English corpus data of (7). Hawkins (1999b) examined not only syntactic weight and the relative sizes of phrasal combination domains, but also various relations of semantic dependency holding between the (intransitive) verb and the PPs. Two entailment tests were conducted. The first, the Verb Entailment Test, asked: does V PP1 PP2 entail V alone, or does V have a meaning dependent on either PP1 or PP2? For example, 'the man waited for his son in the rain' entails 'the man waited'; but 'the man counted on his son in old age' does not entail 'the man counted'. The second, the Pro-Verb Entailment Test, asked: can V be replaced by some general Pro-verb, or does one of the PPs require that exact V for its interpretation. Thus, 'the girl played on the playground' entails 'the girl did something on the playground'; but 'the girl depended on her mother' does not entail 'the girl did something on her mother'. When there was a dependency between V and one of the PPs by these tests, then 73% (151/206) had the dependent PP (PP_d) adjacent to V, i.e. their processing domains were minimal. 82% (265/323) of the sequences in (7) had short PP before long PP, i.e. minimal PCDs. For PPs that were BOTH shorter AND semantically dependent, the adjacency to V was even higher, at 96% (102/106). The combined adjacency effect exerted by semantic dependency and EIC was statistically significantly higher than for each factor alone.

I conclude that the more syntactic and semantic processing operations whose domains can be minimized in a given order, the greater is the preference for that order since each combinatorial and dependency relation can

then be processed faster and with fewer simultaneous demands on working memory.⁶

The full figures illustrating this interaction are given in (15), in which P_d is the PP semantically dependent on V by the Verb or Pro-Verb entailment tests and P_i is the semantically independent PP by these tests:⁷

| (15) | $P_d > P_i$ by | | | $P_d = P_i$ | $P_i > P_d$ | | |
|------------------|----------------|----------|----------|-------------|-------------|----------|-----------|
| n=206 | 5+ | 2-4 | 1 | 1 | 2-4 | 5+ | |
| [V P_d P_i] | 7% (2) | 33% (6) | 74% (17) | 83% (24) | 92% (23) | 96% (49) | 100% (30) |
| [V P_i P_d] | 93% (28) | 67% (13) | 26% (6) | 17% (5) | 8% (2) | 4% (2) | 0% (0) |

I would argue further that these entailment tests provide empirical evidence in favor of assigning COMPLEMENT rather than ADJUNCT status to the P_d . If the relevant V or P cannot be semantically interpreted independently of the other, this would appear to constitute a sufficient criterion for complementhood and for the listing of the P_d within the lexical co-occurrence frame for the V. These data therefore reveal a preference for complements before adjuncts in performance, and an almost exceptionless preference for short complements before adjuncts (and before longer complements in sequences of two complements).

Complements also prefer adjacency to their heads in the basic grammatical orders of numerous phrases in English and other languages and are generated in a position closer to the head than adjuncts in the grammars of

⁶ There is an interesting processing difference between lexical-semantic dependencies like *the man counted on his son in old age* (in which the verb *counted* is dependent on *on his son* for the assignment of its appropriate meaning, but not on *in old age*) and the phrasal combination data and EIC effects. Minimizing the domains for processing the latter can be argued to benefit both the speaker and the hearer (Hawkins 1998a, in preparation), since the relevant (phrasal combination) domains are defined on the basis of the construction or projection potential of terminal elements, and these constructions/projections can define the same domain minimizations are being produced in an on-line fashion (within the Formulator of a production model such as Levelt 1989) or comprehended within a comprehension model. For the lexical-semantic dependencies, on the other hand, the speaker is at an advantage since he/she knows the intended sense of *count* in advance, and yet production data are still highly sensitive to the minimal domains on the basis of which the hearer can disambiguate!

⁷ In corresponding tables cited in Hawkins (1999b, 2001) I included five additional sequences, making 211 in all, in which both the PPs involved an interdependency with V by the entailment tests, but one involved more dependencies than the other. I have excluded these five here, making a total of 206 sequences, in all of which one PP is completely independent of V by both entailment tests discussed in the main text, while the other PP is interdependent with V by at least one entailment test.

Jackendoff (1977) and Pollard and Sag (1987). Tomlin's (1986) Verb Object Bonding principle provides further cross-linguistic support for this by pointing to languages in which it is impossible or dispreferred for adjuncts to intervene between a verbal head and its subcategorized direct object complement. The Adjacency Hypothesis of (14) provides an explanation for this. There are more combinatorial and dependency relations linking complements to their heads whose processing domains can be minimized under adjacency, compared to the relations linking adjuncts and heads. Adjacency of complements is predicted in grammars, therefore, for the same reason that more interdependent and combinatorial processing results in tighter adjacency between V and PP_J in English performance.

Complements are listed in a lexical co-occurrence frame defined by, and activated by, a specific head (e.g. a verb). Adjuncts are not so listed and occur in a wide variety of phrases with which they are semantically compatible. Processing the lexical co-occurrence frame of a verb therefore favors a minimal surface domain in which the verb can be linked to its complements as rapidly as possible.

There are also more productive relations of semantic and syntactic interdependency between heads and complements than between heads and adjuncts, i.e. more cases in which the meaning or grammar of one of these categories requires access to the other for its assignment. E.g. a direct object receives a theta-role from the verb (typically a subtype of Dowty's (1991) Proto-Patient theta-role); adjuncts do not receive theta-roles. The direct object is also syntactically required by a transitive verb; adjuncts are not syntactically required sisters. A transitive verb in turn regularly undergoes what we might call FUNCTION CATEGORY RANGE REDUCTION (following Keenan 1979) by a direct object NP. Compare the different interpretations of *run* in *run the advertisement/run the water/run the race (in the afternoon)*. The verb is lexically dependent on the choice of direct object for its interpretation here, not on an adjunct phrase such as *in the afternoon*. Similarly an intransitive verb is frequently dependent on a PP complement for range reduction (e.g. *count on X*). Adjuncts may, however, be lexically dependent on a verb by Keenan's principle, as function categories (compare the different interpretations of *smoothly* in *sailing smoothly/landing smoothly/talking smoothly*), but the number of semantic and syntactic interdependencies linking them to their heads are fewer, since adjuncts do not receive theta-roles, are not grammatically required, and their heads are not lexically dependent on them.

To these considerations must be added EIC effects. Direct object complements are significantly shorter on aggregate than PP adjuncts or complements (Hawkins 1994), hence their relative adjacency to the verb is predicted for constituent recognition. In fact, given that adjuncts are less inter-

dependent with their heads than complements, the current theory predicts that they will only be adjacent to their heads if they are shorter than accompanying complements, thereby minimizing PCDs.

I therefore see adjacency as motivated by domain minimizations in syntactic and semantic processing. In this way the surface domains in which phrase structure recognition, lexical-semantic dependencies, theta-role assignments and complement relations can be processed will each be smaller and more efficient and will involve less simultaneous processing and less demands on working memory as each relation is processed.

2.3 Domain Minimization in Filler-Gap Domains

When a dependency exists between a c-commanding category A and an asymmetrically c-commanded category B dominated by a sister to A, namely C, then categories A and C can regularly be adjacent, but the precise positioning of the dependent B will depend on its role within category C. Consider the relation that holds between the nominal head of a relative clause and the gap, subcategorizer or resumptive pronoun within the clause that is co-indexed with the head. I have argued in Hawkins (1999a) that numerous hierarchies can be set up on the basis of the increasing domain sizes for relative clause processing, measured in terms of the number of nodes and structural relations that need to be computed in order to match the relative clause head with the co-indexed gap, subcategorizer or resumptive pronoun. One of these is the original Keenan and Comrie (1977) Accessibility Hierarchy, which is formulated as (16) in Comrie (1989) (SU=subject, DO=direct object, IO=indirect object, OBL=oblique, GEN=genitive):

$$(16) \text{SU} > \text{DO} > \text{IO/OBL} > \text{GEN}$$

Examples of relative clauses formed on each of these positions are given in (17):

- | | | |
|---------|--|---------------|
| (17) a. | the professor _i [that O _i wrote the letter] | <i>SU</i> |
| b. | the professor _i [that the student knows O _i] | <i>DO</i> |
| c. | the professor _i [that the student showed the book to O _i] | <i>IO/OBL</i> |
| d. | the professor _i [that the student knows his _i son] | <i>GEN</i> |

Some performance support for this hierarchy as a complexity ranking comes from a repetition experiment conducted by Keenan and Hawkins (1987) on English speakers, children and adults. The prediction was that repetition accuracy would correlate with positions on the hierarchy, subjects being easiest. The data, shown in (18), bear this out (GEN-SU stands for relativ-

zation on a genitive within a subject, GEN-DO for relativization on a genitive within a direct object as in (17d)):

| | | | | | | |
|----------|---|-------|-----|-----|--------|--------|
| (18) | Repetition accuracies for relativization in English | | | | | |
| | SU | DO | IO | OBL | GEN-SU | GEN-DO |
| Adults | 64% | 62.5% | 57% | 52% | 31% | 36% |
| Children | 63% | 51% | 50% | 35% | 21% | 18% |

The relative ranking SU > DO has been corroborated by a number of further studies in the psycholinguistic literature. For example, Pickering and Shillcock (1992) found significant reaction time differences between the two positions in a self-paced reading experiment (cf. Hawkins 1999a for further references).

As the surface domains grow that must be processed in order to link the relative clause head with the position relativized on, the amount of simultaneous processing and the demands on working memory increase. If the position relativized on is a gap, then the very identification of this position is difficult and requires access to the gap's subcategorizer and/or its structural environment, and to the filler (i.e. the relative clause head) upon which the gap is dependent. All of these considerations are reflected in the definition of a Filler-Gap Domain given in Hawkins (1999a), which identifies the smallest amount of surface structure containing information sufficient for the unambiguous parsing of a filler-gap dependency.⁸ Hawkins (1999a) formulates the following general prediction for cross-linguistic variation in relative clauses with Filler-Gap domains:

(19) RELATIVE CLAUSE GAP HIERARCHY

If a relative clause gap is grammatical in position P on a complexity hierarchy H, then gaps will be grammatical in all higher positions on H.

Applied to the Accessibility Hierarchy, this prediction turns out to be exceptionlessly supported. Keenan and Comrie's relative clauses with [-Case] strategies are those containing no nominal element to indicate which posi-

⁸ A FILLER-GAP DOMAIN (FGD) is defined as follows in Hawkins (1999a): An FGD consists of the smallest set of terminal and non-terminal nodes dominated by the mother of a filler and on a connected path that must be accessed for gap identification and processing; for subcategorized gaps the path connects the filler to the gap's subcategorizer and includes, or is extended to include, the gap's dependent and disambiguating arguments (if any); for nonsubcategorized gaps the path connects the filler to the gap site; all constituency relations and cooccurrence requirements holding between these nodes belong to the description of the FGD.

tion is being relativized on and are Filler-Gap structures, therefore. All the languages with [-Case] in their sample (40 in all) limit this gap strategy to all and only the following hierarchically arranged possibilities (sample languages only given here):

(20)[-CASE] GAP LANGUAGES in Keenan and Comrie's Data

| | |
|-------------------------|--|
| SU only: | e.g. Arabic (Classical), Toba Batak, Malagasy, Minang-Kabau |
| SU & DO only: | e.g. Chinese (Pekingese), Fulani, Hebrew, Malay |
| SU & DO & IO/OBL only: | e.g. Basque, Korean, Tamil |
| SU & DO & IO/OBL & GEN: | Japanese |

Hierarchy prediction (19) is a particular instance of prediction (2b) defined by the Performance-Grammar Correspondence Hypothesis of (1). Further complexity rankings of relevance to Filler-Gap relativizations discussed in Hawkins (1999a) are: gaps in CNPC (Complex NP Constraint) environments are more complex than those in non-CNPC environments (see Ross 1967); WH-island environments are more complex than non-WH-island environments; environments with semantically rich bridging verbs or complementizers are more complex than those with semantically weaker items; and others. The cut-off points between grammatical and ungrammatical Filler-Gap relatives across languages are structured by hierarchies of complexity that include these ranked positions.

3 Principle 2: MINIMIZE FORMS

The second principle of efficiency and complexity to be proposed here is (21):

(21) MINIMIZE FORMS

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 expanding the compatibility of *F* with a larger set of properties {*P*}. These minimizations apply in proportion to the ease with which a given *P1* can be assigned in processing to a formally reduced *F* with expanded property compatibilities.

The processing of linguistic forms requires effort. This effort is minimized, first, by reducing the set of formal units in a form or structure. It is minimized, secondly, by reducing the number of forms with unique property assignments. It is not efficient to have a distinct form F for every possible property P that one might wish to express. To do so would greatly increase the number of form-property pairs in a language and the length and complexity of each proposition. Choices have to be made over which properties get priority for unique assignment to forms, and the remaining properties are assigned to forms that are ambiguous, vague or zero-specified with respect to the property in question. It is up to the *context*, broadly construed, to permit assignment of the intended P1 to a form F that is compatible with a larger set of properties {P}. (21) asserts that there is a trade-off between form minimization as defined here and the ease with which such additional properties can be assigned to forms through processes that are variously described as processing enrichments, inferences, implicatures, and sentence-internal dependencies of various sorts (e.g. filler-gap dependencies).

Examples abound whose patterning suggests that a reduction in form processing is an advantage, as long as the relevant information can be recovered, from discourse, from real-world knowledge, or from some accessible linguistic structure. Consider the use of pronouns versus full NPs (*he/she* versus *the professor*), Zipfian (1949) effects (the shorter *TV* for *television*), compounds (*paper plate* for *plate made of paper*; *paper factory* for *factory that makes paper*; *paper clip* for *clip for use on paper*; etc), coordinate deletions (*John cooked 0 and Fred ate the pizza*), and control structures involving understood subjects of verbs within non-finite subordinate clauses (versus corresponding finite clauses with overt subjects). Filler-gap dependencies in e.g. relative clauses are also plausibly motivated by (21). Gaps can be identified by reference to their subcategorizer and to the filler with which they are co-indexed, but the resulting structure is a difficult one for processing and this difficulty is plausibly motivated by the compensating advantage of more minimal form processing (compared with a structure in which the filler is explicitly copied within the relative).

(21) is further supported by the ECONOMY PRINCIPLE of Haiman (1983) and by the data that 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, cf. Levinson *ibid.*).

3.1 Form Minimization Predictions

Minimizing forms and assigning properties through processing enrichments reduces overall processing effort in these cases. But one cannot minimize everything and assign all properties through enrichment. Otherwise human languages would be reduced to a few highly polysemous grunts, beyond the capacity of the processing mechanisms to enrich to the intended P1. There has to be a balance between conventionalized forms and their linguistic properties on the one hand, and on-line inference on the other, and the need for such a balance leads to the following predictions:

(22) 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 P1 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.

In effect, form minimizations require compensating mechanisms. (22a) asserts that frequency and processing ease regulate reductions in form, while frequency and preferred expressiveness regulate the grammaticalization and lexicalization preferences of (22b).

The factors that facilitate form reductions and processing enrichment include high accessibility for the entities and events in the current discourse (whence the use of shorter pronouns and anaphoric VPs rather than full NPs and VPs, cf. Ariel 1990), and real-world knowledge (whence the structurally reduced compounds like *paper plate* in lieu of more explicit post-modifiers such as *made of paper* precisely when the semantics of the post-modifier can be readily inferred, cf. Sperber and Wilson 1995, Levinson 2000). High frequency is also important. Zipf (1949) proposed an inverse correlation between word size and frequency of use in English: short words are more frequent than longer ones in general; and as lexical meanings gain in frequency of usage their formal expression often becomes shorter (whence *TV* for *television*). In grammars, high-frequency closed-class grammatical categories are typically more minimal than their lower-frequency counterparts. Auxiliary verbs are often reduced and contracted in comparison with lexical verbs, each of which has a frequency that is significantly less than that of each auxiliary. Pronouns are generally shorter than full NPs containing lexical nouns, each of which will be less frequent than a grammaticalized pronoun. When definite and indefinite articles emerge historically from

demonstratives and from the numeral *one* respectively, the greater frequency of the articles is matched by a reduction in phonological and/or morphological structure (*the* versus *that*, *a(n)* versus *one* in English, etc). And so on.

Why should frequency be so predictive? This follows, I believe, from the efficiency logic of MINIMIZE FORMS (21). By assigning reduced forms to high-frequency properties, and by reserving more complex forms for lower frequency properties, the overall number of formal units that the processor has to process, in a given sentence and in a lifetime of sentence usage, is less. By contrast, if reduced forms systematically conveyed low-frequency properties and more complex forms high-frequency properties, or if reduced forms were neutral to frequency with the result that some reduced forms expressed frequently used properties while others did not, then more processing of formal units would be required in performance overall. Frequency effects in form reduction therefore accomplish the same minimization in articulation and processing that property expansions and the exploitation of contextual clues do.

Frequency is also relevant to (22b). There are numerous semantic and syntactic properties that are frequently occurring in performance and that have priority in grammatical and lexical conventions across languages. The property of causation is invoked often in everyday language use and is regularly conventionalized in the morphology, syntax or lexicon (Comrie 1989, Shibatani 1976). Agenthood and patienthood are frequently expressed and are 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 baptizing or bequeathing are assigned separate lexical items, but not a uniquely distinctive construction in the syntax (Saddock and Zwicky 1985). Within the lexicon the property associated with *teacher* is frequently used in performance, that of *teacher who is late for class* much less so. The event of *X hitting Y* is frequently selected, that of *X hitting Y with X's left hand* less so. The more frequently selected properties are conventionalized in single lexemes or unique categories, phrases 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 a process of semantic composition.

It is this last point that holds the key to the explanatory question that is raised in connection with (22b): why are frequently used properties given priority when assigning unique properties to forms? The answer is plausibly the same as we gave for form reductions and Zipfian effects. Sensitivity to high frequency results in more minimal forms within a sentence and in a lifetime of sentence usage. If less frequently used properties were given

priority for unique F:P1 assignments, or if these assignments were neutral to frequency with the result that some were high frequency and some were low frequency properties, then more word and phrase combinations would be needed overall whenever the enrichment to P1 was not contextually guaranteed. If *teacher* meant *teacher who is late for class*, then the meaning of *teacher* would have to be expressed by semantic composition through appropriate word and phrase combinations (e.g. by *someone who gives instruction at school level*) since the closest single word would be pre-empted by its low-frequency property assignment. If the relatively infrequent dual number were the base form for a grammatical number system, then the much more frequent singularity and plurality concepts would need to be expressed by longer and more complex combinations (two minus one, two and more, etc). The conventionalization of high-frequency properties in unique F:P1 pairings minimizes forms, therefore.

3.2 Greenberg's Feature Hierarchies

One of the best documented and cross-linguistically researched set of data that test, and support, (22a) and (22b) comes from the FEATURE HIERARCHIES or MARKEDNESS HIERARCHIES of Greenberg (1966) and Croft (1990). Examples are given in (23):

- (23) Nominative > Accusative > Dative > Other (for case marking)
 Singular > Plural > Dual > Trial/Paucal (for number)
 Masculine/Feminine > Neuter (for gender)
 Positive > Comparative > Superlative

Greenberg argued that these hierarchies define relative frequency rankings for the relevant properties in each domain. For example, the relative frequencies of number inflections on nouns in a corpus of Sanskrit are:

- (24) Singular = 70.3%; Plural = 25.1%; Dual = 4.6%.

The other hierarchies have similar frequency correlates. In other words, these hierarchies are PERFORMANCE FREQUENCY RANKINGS defined on entities within common grammatical and/or semantic domains. These performance rankings are then reflected in cross-linguistic grammatical patterns that conventionalize morphosyntax and allomorphy in accordance with the predictions of (22).

One prediction that is made is (25):

(25) QUANTITATIVE FORMAL MARKING PREDICTION

For each hierarchy H the amount of formal marking (i.e. phonological and morphological complexity) will be greater or equal down each hierarchy position.

(25) follows from (22a). For example, in Manam the 3rd Singular suffix on nouns is zero, the 3rd Plural is *-di*, the 3rd Dual is *-di-a-ru* and the 3rd Paucal is *-di-a-to* (Lichtenberk 1983). The amount of formal marking increases from singular to plural, and from plural to dual, and is equal from dual to paucal, in accordance with the hierarchy in (23). Similarly English singular nouns are zero-marked whereas plurals are formally marked, generally with an *-s* allomorph.

(26) MORPHOLOGICAL INVENTORY PREDICTION

For each hierarchy H ($A > B > C$) if a language assigns at least one morpheme uniquely to C, then it assigns at least one uniquely to B; and if it assigns at least one uniquely to B, it does so to A.

(26) follows from (22b). A distinct Dual implies a distinct Plural and Singular in the grammar of Sanskrit, and a distinct Dative implies a distinct Accusative and Nominative in the case grammar of Latin and German (or a distinct Ergative and Absolutive in Basque, cf. Primus 1999). A unique number or case assignment low in the hierarchy implies unique and differentiated numbers and cases in all higher positions.

(27) DECLINING DISTINCTIONS PREDICTION

For each hierarchy H any combinatorial features that partition references to a given position on H will result in fewer or equal morphological distinctions down each lower position of H.

(27) also follows from (22b). For example, unique and gender-distinctive pronouns can exist for the singular and not for the plural in English (*he/she/it* versus *they*), whereas the converse uniqueness is not predicted.

I have argued in Hawkins (1998b, c) that we should view the performance frequencies of different hierarchy positions not just as correlations of these hierarchies, but as causes of the grammatical asymmetries in the conventionalization of morpho-syntactic distinctions, in accordance with (22).

These grammatical distinctions and their degrees of formal marking reflect usage: the more frequently a form can be used to identify some property in performance, the more it will be grammaticalized, and the more reduced its surface marking can be. Singular entities are referenced more frequently in performance than plurals and duals, nominatives more frequently than accusatives and datives, and these frequencies are then reflected in the conventionalized preferences of grammars, favoring more grammatical distinctions but less explicit marking in the higher positions compared to the lower ones. The ultimate causes of the frequencies, in turn, can be diverse and can reflect real-world frequencies of occurrence (pairs are less frequent than singular entities and undifferentiated plurals) and communicative biases (in favor of animates rather than inanimates, cf. gender systems). The frequencies of use for the different hierarchy positions can also reflect degrees of syntactic and semantic complexity predicted by other efficiency principles (which is plausibly what we see in the declining use of the lower case forms in increasingly complex clauses with more argument positions).

All of these data exemplify the testing of prediction (2b) of the PERFORMANCE GRAMMAR CORRESPONDENCE HYPOTHESIS in section 1.

3.3 Form Minimization and Adjacency

There is an interesting interaction between MINIMIZE FORMS (21) and MINIMIZE DOMAINS (3) that can be seen both in performance and in correlating grammatical data across languages. When one structure involves fewer overt forms than an alternative, then the syntactic or semantic properties that are conveyed by the missing forms must be recovered from their environments. As a result there are more relations of syntactic or semantic dependency between categories, whose processing assigns the properties in question, while fewer properties can be derived from these categories independently of their environments. There are also often longer domains for processing these properties when overt forms are absent. As a consequence, constructions with fewer forms define a stronger preference for minimal domains of processing compared with their explicitly marked counterparts, for the same basic reason that additional semantic dependencies result in tighter adjacency in the prepositional phrase data of section 2.2.

Consider the alternation in English between WH or THAT relativizers versus ZERO in restrictive relative clauses. ZERO relatives, e.g. *the Danes you taught*, involve a clause (*you taught*) that requires more dependent processing on the head. For example, processing the subcategorization frame and the precise semantics of this subordinate verb requires access to the subject *you* and to the second (direct object) argument. If that argument is represented by a WH pronoun (*the Danes whom you taught*) the appro-

appropriate properties can be assigned to *taught* on the basis of material within the relative clause itself (as happens also in pronoun-retaining languages like Hebrew, see below). But in the absence of WH, the processing of *taught* must access the head, *Danes*, and this introduces a dependency on the head whose processing prefers a minimal domain. Processing the co-indexation relation between the head and the relative also requires direct access to the subcategorizer *taught* (or to its gap position, depending on one's theory) since there is no intervening WH word to bear the index. As a result of these and additional processing dependencies that are defined explicitly in Hawkins (in preparation), tighter adjacency is predicted between ZERO relatives and their heads than between WH and THAT relatives and their heads. The following data are from Quirk (1957):

- (28) a. Restrictive (non-subject) relatives adjacent to the head
 WH = 28% (152) THAT = 32% (174) ZERO = 41% (222)
 b. Restrictive (non-subject) relatives with intervening material
 WH = 50% (31) THAT = 44% (27) ZERO = 7% (4)

Only 4/226 (2%) of the more dependent ZERO relatives are non-adjacent to the head.

Across languages grammatical rules have conventionalized such preferences of performance: zero-marked dependencies require adjacency to their heads, formally marked dependencies occur in non-adjacent and adjacent domains. Consider the following agreement universal for languages with case copying, from Moravcsik (1995):

- (29) If agreement through case copying applies to NP constituents that are adjacent, it applies to those that are non-adjacent.

For example, case copying is quite systematic in the Australian language Kalkatungu (a word-marking language according to Blake 1987) and occurs when NP constituents are both adjacent and non-adjacent:

- (30) a. thuku-yu yant-tu yanyi itya-mi
 dog-ERG big-ERG white-man bite-FUT
 b. thuku-yu yanyi itya-mi yaun-tu
 dog-ERG white-man bite-FUT big-ERG

But in Warlpiri (a phrase-marking language) case copying is found only on non-adjacent NP constituents, cf. (31b), while adjacent sisters (cf. (31a)) have only one occurrence of case marking in conjunction with zero-marking:

- (31)a. tyarntu wiri-ngki+tyu yarlki-rnu
 dog big-ERG +me bite-PAST
- b. tyarntu-ngku+tyu yarlku-rnu wiri-ngki
 dog-ERG +me bite-PAST big-ERG

Similarly in highly inflected languages like Latin, separation of Adj and N is made possible by rich morphological agreement (see Vincent 1987), suggesting more generally that if a language permits separation of N and Adj, then Adj must agree with N.

The grammatical variation here reveals a conventionalization of exactly the preferences that are evident in English performance data such as (28). Under conditions of adjacency both formal marking and zero have been grammatically fixed, the former in Kalkatungu and Latin, the latter in Warlpiri, and these two options are both productive in performance, cf. (28a). But under conditions of non-adjacency only formal marking is productive in grammars, and only this option is productive in performance as well, cf. (28b).

3.4 Form Minimization in Filler-Gap Domains

Similarly the gap strategy in relativization prefers a processing domain that is smaller and simpler than that of relative clauses with explicit marking of the position relativized on. Explicit marking can be achieved by using a copy pronoun, as in Hebrew:

- (32) ha-ishaj [she-Yon natan laj et ha-sefer]
 the woman that John gave to-her DO the book
 'the woman that John gave the book to'

Copy pronouns make identification of the position relativized on as easy as it can possibly be. They also permit the processing of the verb's subcategorization and semantics to take place locally in the relative clause and without having to access the head noun, which the gap-containing structures must do, as in the English ZERO relative construction. This leads to the prediction that gaps will prefer more minimal domains for processing, while copy pronouns fill the more complex positions on hierarchies such as the Accessibility Hierarchy, in proportion to the degree of complexity involved. We therefore predict that if a copy pronoun is grammatical in position P on a complexity hierarchy H, then it will be grammatical in all lower positions that can be relativized at all (cf. Hawkins 1999a). 27 of Keenan and Comrie's (1977) languages have copy pronouns (as a subinstance of their

[+Case] relativization strategy), and this prediction holds exceptionlessly for these languages (sample languages only given here):

(33) Pronoun-retaining Languages in Keenan and Comrie's Data

| | |
|-------------------------|--|
| GEN only: | e.g. Japanese, Javanese, Malay, Turkish |
| GEN & IO/OBL only: | e.g. Toba Batak, Hausa, Welsh |
| GEN & IO/OBL & DO only: | e.g. Chinese (Pekingese), Kera, Slovenian |
| GEN & IO/OBL & DO & SU: | Urhobo |

The grammatical data from relative clauses reveal two opposite implicational patterns overall: gaps go from low to high on a complexity hierarchy (see (20)), copy pronouns from high to low. In Hawkins (*ibid.*) I quantified the distribution of gaps to copy pronouns in 24 languages (like Hebrew) that have both: the occurrence of gaps declined down the Accessibility Hierarchy of (16) by 100% to 65% to 25% to 4%, while that of pronouns increased (0% to 35% to 75% to 96%). The gapped relatives involve more dependencies between relative and head, therefore, and the attested Filler-Gap Domains are smaller than corresponding Filler-Copy Pronoun domains in languages like Hebrew, in accordance with Minimize Domains (3).

4 Principle 3: MAXIMIZE ON-LINE PROCESSING

The third principle I propose is (34):

(34) MAXIMIZE ON-LINE PROCESSING

The human parser prefers to maximize the set of properties that are assignable to each item X as X is parsed. The maximization difference between competing orders and structures will be a function of the number of properties that are misassigned or unassigned to X in a structure/sequence S, compared with the number in an alternative.

This principle asserts that it is preferable to be able to recognize syntactic and semantic properties productively throughout the processing of a sentence.⁹ What is dispreferred is, first, any significant delay or LOOK AHEAD

⁹ Notice that Maximize On-line Processing is formulated in terms of parsing and the hearer, since the speaker does not make structural misassignments on-line and can enrich unassignments based on knowledge of what is to be produced later. Nonetheless there is evidence for the speaker's accommodation to the hearer's needs here, in proportion to the degree of misas-

(Marcus 1980) in on-line property assignments, and second a misassignment of properties on-line. Misassignments result in so-called garden path effects whereby one analysis is chosen on-line and is then subsequently corrected in favor of a different analysis when more material has been processed. A famous example is *the horse raced past the barn fell* which is first assigned a main clause reading and then a reduced relative reading when the (matrix verb) *fell* is encountered (see MacDonald et. al. 1994). Such backtracking is difficult for the processor, but more importantly it is inefficient since initial property assignments are wasted and make no contribution to the ultimate syntactic and semantic representation of the sentence. Similarly any processing delay followed by a late surge of property assignments towards the end of a clause or sentence does not maximize on-line property assignments.

This principle is discussed and exemplified in detail in Hawkins (in press-ab). I believe that it is ultimately responsible for a large number of linear precedence asymmetries across languages which contrast with the kinds of symmetrical orderings associated with the head-initial and head-final typology illustrated in (10) and (11) above. These asymmetries involve a clear preference for just one ordering of categories across languages. Many involve cases in which there is an asymmetric dependency between some category D and another category I.

4.1 Fillers First

For example, gaps are dependent on their fillers in Filler-Gap structures for coreference and co-indexation and also for recognizing the position to be filled (in conjunction with access to the subcategorizer, if there is one), whereas fillers are not so dependent on their gaps. This results in a preference for fillers before gaps, or Fillers First (Hawkins 1999a, J.D. Fodor 1983). When the gap follows the filler, I ... D, then the filler (I) can be fully processed on-line, and the grammatical properties that are assigned by reference to I can be assigned to the gap on-line as well, resulting in an efficient distribution of property assignments throughout the sentence. But if the gap precedes, D ... I, its full properties can only be assigned retrospectively when the filler is encountered, resulting in a processing delay and in frequent garden path effects as matrix and subordinate clause arguments are redistributed to take account of the gap that is activated by late processing of the filler (cf. Antinucci et. al. 1979, Clancy et. al. 1986). I ... D maximizes on-line property assignments, therefore.

When the filler is a WH-word in a WH-question there is unambiguous cross-linguistic support for Fillers First: almost all languages that move a

signment and unassignment involved in various structures, and a quantified metric for testing Maximize Online Processing is proposed in Hawkins (2002a).

WH-word to clause peripheral position move it to the left, not to the right (Hawkins 1999a: in press-ab). In relative clauses there is also clear support, but Fillers First is now in partial conflict with the EIC preference for noun-final NPs in head-final languages. Head-initial languages have consistently right-branching relatives (e.g. [V [N S]]), which are motivated both by EIC and by Fillers First. But head-final languages have either left-branching relatives ([[S N] V]), which is good for EIC but which positions the gap before the filler, or right-branching relatives ([[N S] V]), which is good for Fillers First but which creates non-adjacency between heads and makes PCDs longer.

The variation here points to the existence of two preferences, whose predictions overlap in one language type but conflict in another. This illustrates prediction (2c) of the PERFORMANCE-GRAMMAR CORRESPONDENCE HYPOTHESIS in (1). Each of the structural variants is motivated. There is also suggestive evidence that each principle is exemplified across grammars in proportion to its own degree of motivation. The head-final languages that prefer left-branching relatives are the *rigid* ones like Japanese, in which there are more containing V-final and P-final phrases, etc, that prefer the head of NP to be final as well (by EIC). *Non-rigid* head-final languages have fewer containing phrases that are head-final and so define a weaker preference for noun-finality, allowing Fillers First to assert itself more, which results in more right-branching relatives (cf. Lehmann 1984 for numerous exemplifying languages).

4.2 Other Linear Precedence Asymmetries

A related structure involves topicalized XPs with gaps in a sister S. These generally precede S across languages (Gundel 1974, 1988; Primus 1993, 1999). The reverse ordering could be optimal both for EIC and for scope marking, but it is either ungrammatical or dispreferred and this provides further evidence for Fillers First. The asymmetry disappears when a co-indexed pronoun replaces the gap, resulting in left- or right-dislocation structures, suggesting that it is the gap that contributes substantially to the linear precedence asymmetry.

The dispreference for D ... I here is motivated by the dependence of D (the gap) on I (the filler) for gap identification and filling, as before. In addition the ABOUTNESS relation between the predication and the topic (Reinhart 1981), coupled with the regular referential independence or givenness of the topic, means that the semantic processing of the predication is often incomplete without prior access to the topic (cf. Keenan 1978, Gundel 1988), whereas the topic can be processed independently of the predication. For example, Tsao (1978) gives numerous examples from Mandarin

Chinese of the topic phrase providing information that is required for interpretation of the predication, making these predications *dependent* on the topic as this term is defined here. The topic can provide the understood possessor for a possessee within the predication, a class or set relative to which some entity in the predication is understood to be a member, a domain relative to which a superlative claim in the predication is interpreted, a restrictive adjunct for a noun in the predication, a locative or time adjunct for a verb, and so on. In all these cases the proper interpretation of the predication depends on the topic and these asymmetric dependencies motivate an asymmetric topic before predication ordering across languages, I ... D, thereby maximizing on-line processing efficiency (Hawkins 2002a).

Further ordering asymmetries that are plausibly motivated by (34) include the preference for wide scope before narrow scope operators and quantifiers, and also the preference for restrictive before appositive relatives exemplified in the English (35a):

- (35) a. Students that major in mathematics, who must of course work hard, ... R + A
 b. *Students, who must of course work hard, that major in mathematics, ... A + R

In the on-line processing of (35b) there would always be a semantic garden path. The appositive relative would first be predicated of all students, and would then be predicated only of that subset of which the restrictive relative was true, once the latter was encountered and processed. The ordering of (35a) avoids the regular garden path by placing together all the items that determine the reference set of which the appositive clause is predicated, and by positioning them before an appositive claim that would make an erroneous predication on-line if it preceded the restrictive clause. The positioning of the restrictive is also motivated by the Adjacency Hypothesis in (14), since the nominal head is semantically dependent on the restrictive but not on the appositive clause for the restriction in its reference. Principles (34) and (14) therefore reinforce one another in this case.

5 Conclusions

I conclude that there is a profound correspondence between performance data and grammars, see the Performance-Grammar Correspondence Hypothesis of (1) and (2). Structural types that are preferred in performance are also preferred in grammars (2a), preference rankings in performance are matched by hierarchies of grammatical conventions (2b), and partial competition between preferences results in grammatical variation (2c).

Three general principles of efficiency and complexity have been proposed: MINIMIZE DOMAINS (3); MINIMIZE FORMS (21); and Maximize ON-LINE PROCESSING (34). These principles, individually and in combination, can motivate a broad range of preference data in performance and in grammars. They are simple and intuitive principles that reflect an even more general Zipfian principle of least effort (cf. Zipf 1949), yet they can explain many subtle properties of syntax that have been largely viewed as innate and non-functional hitherto, in accordance with Chomsky (1965) and subsequent publications. Yet when precise functional principles are formulated, I would argue that they provide an alternative to this view, with wide-ranging empirical support, less stipulation, and more independently motivated explanation.

It remains to be seen how much of core syntax and of syntactic variation can be explained in these terms. This general approach views grammars and grammatical evolution as complex adaptive systems (Gell-Mann 1992), with efficiency and ease of processing driving the adaptation in response to prior changes. Innate syntactic knowledge is not the ultimate explanation, although the processing architecture that defines ease of use is most plausibly innate.

It will also be necessary to answer questions that are raised by this general approach. How exactly do the preferences of performance gradually become fixed conventions in language evolution, whereby only the preferred structure is generated and dispreferred options are eliminated altogether? Kirby (1999) gives a clear discussion of the issues here and provides an intriguing computer simulation of grammars evolving out of performance preferences. Haspelmath (1999) discusses the question from the perspective of Optimality Theory and argues that the constraints of this theory can be functionally motivated by performance preferences such as those proposed here which become grammaticalized through a process of diachronic adaptation. There are also psycholinguistic issues that are raised by these performance preferences. How exactly can the kinds of efficiencies described here be implemented in current production and comprehension models with the result that they could actually be predicted? And to what extent do the needs and benefits of the speaker overlap with those of the hearer, to what extent are they different, and to what extent does the speaker accommodate to the hearer?

These are big issues that arise independently of the central hypothesis of this paper, which is that there is a profound correspondence between the preferences of performance and those of grammars, whatever the precise causality of the performance data turns out to be. This hypothesis is at variance with the proposed autonomy of grammars from performance which has dominated generative thinking since Chomsky (1965). The evidence of this

paper suggests that syntax is, to a significant extent at least, PERFORMANCE-DRIVEN and that we can achieve a richer and more explanatory theory of grammar if we combine generative insights with a theory of processing efficiency of the type outlined here.

References

- Antinucci, F., A. Duranti, and L. Gebert. 1979. Relative Clause Structure, Relative Clause Perception, and the Change from SOV to SVO. *Cognition* 7: 145-176.
- Bever, T.G. 1970. The Cognitive Basis for Linguistic Structures. *Cognition and the Development of Language*, ed. J.R. Hayes, 279-362. New York: Wiley.
- Blake, B. 1987. *Australian Aboriginal Grammar*. London: Routledge.
- Chomsky, N. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- Clancy, P.M., H. Lee, and M. Zoh. 1986. Processing Strategies in the Acquisition of Relative Clauses: Universal Principles and Language-Specific Realizations. *Cognition* 14: 225-262.
- Comrie, B. 1989. *Language Universals and Linguistic Typology*. 2nd ed. Basil Blackwell, Oxford.
- Croft, W. 1990. *Typology and Universals*. Cambridge, Cambridge University Press.
- Dowty, D.R. 1991. Thematic Proto-Roles and Argument Selection. *Linguistics* 67: 547-619.
- Dryer, M.S. 1992. The Greenbergian Word Order Correlations. *Language* 68: 81-138.
- Dryer, M.S. 1993. Modifying Hawkins' Prepositional Noun Modifier Hierarchy. Paper Presented at the Annual Meeting of the Linguistic Society of America, Los Angeles.
- Fodor, Janet D. 1983. Phrase Structure Parsing and the Island Constraints. *Linguistics and Philosophy* 6: 163-223.
- Gell-Mann, M. 1992. Complexity and Complex Adaptive Systems. *The Evolution of Human Languages*, eds. J.A. Hawkins and M. Gell-Mann, 3-18. Redwood City, CA: Addison-Wesley.
- Greenberg, J.H. 1963. Some Universals of Grammar with Particular Reference to the Order of Meaningful Elements. *Universals of Language* 2nd ed., ed. J.H. Greenberg, 73-113. Cambridge, MA: MIT Press.
- Greenberg, J.H. 1966. *Language Universals, with Special Reference to Feature Hierarchies*. The Hague: Mouton.
- Grice, H.P. 1975. Logic and Conversation. *Speech Acts. Syntax and Semantics* vol.3, eds. P. Cole and J. Morgan, 41-58. New York: Academic Press.
- Gundel, J.K. 1974. The Role of Topic and Comment in Linguistic Theory. Doctoral dissertation, University of Texas, Austin.

- Gundel, J.K. 1988. Universals of Topic-Comment Structure. *Studies in Syn-tactic Typology*, eds. M Hammond, E.A. Moravcsik, and J.R. Wirth, 209-239. Amsterdam: John Benjamins.
- Haiman, J. 1983. Iconic and Economic Motivation. *Language* 59: 781-819.
- 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. 1993. Heads, Parsing, and Word Order Universals. *Heads in Gram-matical Theory*, eds. G.G. Corbett, N.M. Fraser, and S. McGlashan, 231-265. Cambridge: Cambridge University Press.
- Hawkins, J.A. 1994. *A Performance Theory of Order and Constituency*. Cambridge: Cambridge University Press.
- Hawkins, J.A. 1998a. Some Issues in a Performance Theory of Word Order. *Con-stituent Order in the Languages of Europe*, ed. A. Siewierska, 729-781. Berlin: de Gruyter.
- Hawkins, J.A. 1998b. Morphological Hierarchies in the History of Ger-manic. *Interdisciplinary Journal for Germanic Linguistics and Semiotic Analysis* 3: 197-217.
- Hawkins, J.A. 1998c. A Typological Approach to Germanic Morphology. *His-torische Germanische und Deutsche Syntax*, ed. J.O. Aske-dal, 49-68. Frankfurt: Peter Lang Verlag.
- Hawkins, J.A. 1999a. Processing Complexity and Filler-Gap Dependencies across Grammars. *Language* 75: 244-285.
- Hawkins, J.A. 1999b. The Relative Order of Prepositional Phrases in Eng-lish: 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. 2002a. Symmetries and Asymmetries: Their Grammar, Typology and Parsing. *Theoretical Linguistics* 28: 95-150.
- Hawkins, J.A. 2002b. Issues at the Performance-Grammar Interface: Some Com-ments on the Commentaries. *Theoretical Linguistics* 28: 211-226.
- Hawkins, J.A. in prep. *Efficiency and Complexity in Grammars*. Oxford: Oxford University Press.
- Jackendoff, R. 1977. *X-bar Syntax: A Study of Phrase Structure*. Cam-bridge, MA: MIT Press.
- Keenan, E.L. 1975. Variation in Universal Grammar. *Analyzing Variation in Eng-lish*, eds. R. Fasold and R. Shuy, 136-148. Washington DC: Georgetown Uni-versity Press.
- Keenan, E.L. 1978. The Syntax of Subject-Final Languages. *Syntactic Ty-pology: Studies in the Phenomenology of Language*, ed. W.P. Leh-mann, 267-327. Aus-tin: University of Texas Press.

- Keenan, E.L. 1979. On Surface Form and Logical Form. *Studies in the Linguistic Sciences* 8: 163-203.
- Keenan, E.L. and B. Comrie. 1977. Noun Phrase Accessibility and Universal Grammar. *Linguistic Inquiry* 8: 63-99.
- Keenan, E.L. and S. Hawkins. 1987. The Psychological Validity of the Accessibility Hierarchy. *Universal Grammar: Fifteen Essays*, ed. E.L. Keenan, 60-85. London: Routledge.
- Kimball, J. 1973. Seven Principles of Surface Structure Parsing in Natural Language. *Cognition* 2: 15-47.
- Kirby, S. 1999. *Function, Selection and Innateness: The Emergence of Language Universals*. Oxford: Oxford University Press.
- Lehmann, C. 1984. *Der Relativsatz*. Tübingen: Gunter Narr Verlag.
- Levelt, W.J.M. 1989. *Speaking: From Intention to Articulation*. Cambridge, MA: MIT Press.
- Levinson, S.C. 2000. *Presumptive Meanings: The Theory of Generalized Conversational Implicature*. Cambridge, MA: MIT Press.
- Lichtenberk, F. 1983. *A Grammar of Manam*. Honolulu: University of Hawaii Press.
- MacDonald, M.C., N.J. Pearlmutter, and M.S. Seidenberg. 1994. The Lexical Nature of Syntactic Ambiguity Resolution. *Psychological Review* 101: 676-703.
- Marcus, M.P. 1980. *A Theory of Syntactic Recognition for Natural Language*. Cambridge, MA: MIT Press.
- Miller, G.A. and N. Chomsky. 1963. Finitary Models of Language Users. *Handbook of Mathematical Psychology* vol.2, eds. R.D. Luce, R.R. Bush and E. Galanter, 419-492. New York: Wiley.
- Moravcsik, E.A. 1995. Summing up Suffixaufnahme. *Double Case: Agreement by Suffixaufnahme*, ed. F. Plank, 451-484. Oxford: Oxford University Press.
- Pickering, M. and R. Shillcock. 1992. Processing Subject Extractions. *Island Constraints: Theory, Acquisition and Processing*, eds. H. Goodluck and M. Rochemont. Dordrecht: Kluwer.
- Pollard, C. and I.A. Sag 1987. Information-Based Syntax and Semantics, Volume 1: Fundamentals. *Center for the Study of Language and Information Lecture Notes No.13*. Stanford: CSLI Publications.
- Pollard, C. and I.A. Sag 1994. *Head-Driven Phrase Structure Grammar*. Chicago: University of Chicago Press.
- Primus, B. 1993. Word Order and Information Structure: A Performance-Based Account of Topic Positions and Focus Positions. *Syntax: An International Handbook of Contemporary Research* vol. 1, eds. J. Jacobs et. al., 880-896. Berlin: de Gruyter.
- Primus, B. 1999. *Cases and Thematic Roles: Ergative, Accusative, Active*. Tübingen: Max Niemeyer Verlag.

- Quirk, R. 1957. Relative Clauses in Educated Spoken English. *English Studies* 38: 97-109.
- Reinhart, T. 1981. Pragmatics and Linguistics: An Analysis of Sentence Topics. *Philosophica* 27: 53-94.
- Ross, J.R. 1967. Constraints on Variables in Syntax. Doctoral dissertation, MIT.
- Sadock, J.M. and A.M. Zwicky. 1985. Speech Act Distinctions in Syntax. *Language Typology and Syntactic Description* vol. 1, *Clause Structure*, ed. T. Shopen, 155-196. Cambridge: Cambridge University Press.
- Shibatani, M. ed. 1976. *The Grammar of Causative Constructions*. New York: Academic Press.
- Sperber, D. and D. Wilson. 1995. *Relevance: Communication and Cognition*. 2nd ed. Oxford: Blackwell.
- Stallings, L.M. 1998. Evaluating Heaviness: Relative Weight in the Spoken Production of Heavy-NP Shift. Doctoral dissertation, University of Southern California.
- Tomlin, R.S. 1986. *Basic Word Order: Functional Principles*. London: Croom Helm.
- Tsao, F-F. 1978. *A Functional Study of Topic in Chinese*. Taipei: Student Book Company.
- Vincent, N.B. 1987. Latin. *The Romance Languages*, eds. M.B. Harris and N.B. Vincent, 26-78. Oxford: Oxford University Press.
- Wasow, T. 2002. *Postverbal Behavior*. Stanford: CSLI Publications.
- Yamashita, H. and F. Chang. 2001. 'Long before Short' Preference in the Production of a Head-Final Language. *Cognition* 81: B45-55.
- Zipf, G. 1949. *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology*. New York: Hafner.

