Title

Processing Efficiency and Complexity in Typological Patterns

ISBN

9780199281251

Author

Hawkins, JA

Publication Date

2010-11-25

DOI

10.1093/oxfordhb/9780199281251.013.0012

Peer reviewed

Processing Efficiency and Complexity in Typological Patterns

John A. Hawkins, Cambridge University

0.1 Introduction

This chapter presents a research program in which typological patterns are ultimately explained in terms of language processing and use. The central hypothesis is that grammars (implicational universals, hierarchies and distributional preferences) are conventionalizations of the patterns and preferences that one observes in the performance of languages with structural choices (between competing word orders, relative clause structures, morphological alternatives, etc). A number of typologists have been coming to this conclusion in recent years, and in Hawkins (2004) I refer to it as the "Performance-Grammar Correspondence Hypothesis":

(1) Performance-Grammar Correspondence Hypothesis (PGCH)

Grammars have conventionalized syntactic structures in proportion to their degree of preference in performance, as evidenced by patterns of selection in corpora and by ease of processing in psycholinguistic experiments.

Greenberg (1966) was the first to draw attention to such correlating patterns in his discussion of markedness hierarchies like Singular > Plural > Dual > Trial/Paucal. Morphological inventories across grammars and declining allomorphy provided evidence for these hierarchies, while declining frequencies of use in languages with rich inventories suggested not only a correlation with performance but a possibly causal role for it in the evolution of the grammatical regularities themselves (Greenberg 1995:163-164), cf. §0.3.2 below for illustration. Givón (1979:26-31) meanwhile observed that performance

preferences in one language corresponded to an actual categorical requirement for the relevant rule or property in another. The strong preference for definite over indefinite grammatical subjects in English, for example, has been conventionalized into a categorical requirement for definite subjects in Krio and other languages. Bybee & Hopper (2001) document the clear role of frequency in the emergence of a number of grammatical structures, and in Hawkins (1990, 1994, 2004) I argued that the preferred word orders in languages with choices are those that are most productively conventionalized as fixed orders in languages with less freedom.

The PGCH in (1) defines a very different relationship between performance and grammars than the classic one presented in Chomsky (1965) and subsequent publications. Although the (competence) grammar is an important component of an overall performance model for Chomsky, he has argued that grammars are ultimately autonomous and independent of performance factors and are determined by an innate U(niversal) G(rammar). In order to test the PGCH, therefore, we need to examine variation data both across and within languages. If patterns in the one (in grammars) match patterns in the other (in performance), the hypothesis will be supported. If there is no such match, it will not be. I will argue in this paper that there is significant support for the PGCH. To make the discussion less anecdotal, illustrative data will be presented around some general organizing principles that describe common patterns in grammars and performance. Three of these will be presented here (following Hawkins 2004): Minimize Domains (§0.2), Minimize Forms (§0.3), and Maximize On-line Processing (§0.4). The first will be illustrated with patterns involving relative clauses, the second with morphological data and markedness hierarchies, and the third with a number of linear precedence regularities that hold across different

language types. These principles are not claimed to be exhaustive or exclusive of others, but simply to have wide applicability to a broad range of patterns. Section 0.5 presents my conclusions, summarizes some general issues raised by this approach to linguistic typology, and discusses challenges that remain.

0.2 Minimize Domains

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

(2) Minimize Domains (MiD)

The human processor prefers to minimize the connected sequences of linguistic forms and their conventionally associated syntactic and semantic properties in which relations of combination and/or dependency are processed. The degree of this preference is proportional to the number of relations whose domains can be minimized in competing sequences or structures, and to the extent of the minimization difference in each domain.

Combination = Two categories A and B are in a relation of combination iff they occur within the same syntactic mother phrase or maximal projection (phrasal

combination), or if they occur within the same lexical co-occurrence frame (lexical combination).

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

Consider relative clause formation. It involves a dependency between the head of the relative clause and the position relativized on, i.e. the gap, subcategorizor or resumptive pronoun within the clause that is co-indexed with the head, cf. Hawkins (1999, 2004) for a summary of the different formalizations and theories here. I have argued in Hawkins (op cit) that various hierarchies can be set up on the basis of increasing domain sizes for relative clause processing, measured in terms of the smallest number of nodes and structural relations that must be computed in order to match the relative clause head with the co-indexed gap, subcategorizor or resumptive pronoun. One of these is the original Keenan and Comrie (1977) Accessibility Hierarchy, which is formulated as (3) in Comrie (1989) (SU=subject, DO=direct object, IO=indirect object, OBL=oblique, GEN=genitive):

(3) Accessibility Hierarchy (AH): SU>DO >IO/OBL>GEN

Examples of relative clauses formed on each of these positions are given in (4):

(4) 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*

A "filler-gap" or "filler-subcategorizor" domain for relativization on the DO position necessarily contains a co-occurring SU (and more phrasal nodes), a relative on a SU need not contain (and regularly does not contain) a DO. A relativized IO contains a SU and a DO. It is co-occurrence asymmetries such as these between arguments, coupled with the added phrasal complexity of the lower AH positions (OBL and especially GEN) that, I believe, underlies the Keenan-Comrie hierarchy. Whether this is the correct account or not, there are clear patterns across grammars, and there are equally clear correlating patterns in performance, which I shall now summarize.

0.2.1 Patterns in the grammar of relative clauses

One of the most striking patterns that Keenan & Comrie (1977) presented in favor of (3) involved languages that "cut off" at different points down the hierarchy, i.e. their grammars permitted relative clauses to be formed on all higher positions above the cut-off, but not on lower positions. Illustrative languages cited by Keenan & Comrie are those in (5):

(5) Rules of relative clause formation and their cut-offs within the clause

SU only: Malagasy, Maori

SU & DO only: Kinyarwanda, Indonesian

SU & DO & IO only: Basque

SU & DO & IO & OBL only: North Frisian, Catalan

SU & DO & IO & OBL & GEN: English, Hausa

A further pattern involved the distribution of gap strategies ([-Case] in Keenan & Comrie's terminology) and resumptive pronouns (as a type of [+Case] strategy). The difference between the two can be illustrated with the following pair from Hebrew (Ariel 1990):

(6) a Shoshana hi ha-isha*i* [she-nili ohevet 0*i*]

Shoshana is the-woman that-Nili loves

b Shoshana hi ha-isha*i* [she-nili ohevet ota*i*]

that-Nili loves her

(6a) involves a gap, (6b) a resumptive pronoun. Languages with gaps show the same hierarchy pattern as (5), i.e. for relativization as a whole (regardless of strategy): if a gap is grammatical on a low position of AH, it is grammatical on all higher positions. Resumptive pronouns show the reverse pattern: if a resumptive pronoun is grammatical on a high position, it is grammatical on all lower positions (that can be relativized at all).

This can be seen graphically in Table 1 in which I quantify the distribution of gaps to pronouns for 24 languages from the Keenan-Comrie language sample that have both. Gaps decline down the AH, 100% to 65% to 25% to 4%, pronouns increase (0% to 35% to 75% to 96%).

INSERT TABLE 1 HERE

The intuition that emerges from this reverse hierarchy pattern is that gaps are associated with simpler environments (the smaller "filler-gap domains", especially SU and DO) and extend to lower positions only if all higher AH positions also permit a gap. Conversely pronouns favor more complex environments (GEN and OBL) and extend to simpler ones only if the complex positions also permit a pronoun. A plausible explanation that will be supported by the performance data in §0.2.2 is that gaps are harder to process than resumptive pronouns and prefer smaller structural domains for the various relations that need to be computed in

relative clause processing. For example, the pronoun ota in (6b) provides a local and minimal domain for processing the lexical co-occurrences (i.e. the argument structure) of the verb natan (loves) and does not need to extend this search for arguments to the head of the relative itself (isha). Only co-indexing need apply non-locally linking isha*i* and la*i*, making domains of processing more minimal overall, cf. below.

Numerous language-particular rules confirm this pattern of gaps in smaller relativization domains and pronouns in larger ones, e.g. in Cantonese. The pronoun is ungrammatical in the simple relative (7b) but grammatical in (8), in which there is a bigger distance between co-indexed pronoun and relative clause head, i.e. a more complex relativization domain (Matthews & Yip 2003):

(7) a [Ngo5 ceng2 0*i*] go2 di1 pang4jau5*i*

I invite those CL friend

'friends that I invite'

. b *[Ngo5 ceng2*i* keoi5dei6*i*)] go2 di1 pang4jau5*i*

I invite them those CL friend

(8) [Ngo5 ceng2 (keoi5dei6*i*) sik6-faan6] go2 di1 pang4jau5*i*

I invite (them) eat-rice those CL friend

'friends that I invite to have dinner'

It should be pointed out that these patterns and limitations on relativization are quite surprising from a purely grammatical perspective. They are different from the kinds of subjacency constraints of Ross (1966), Chomsky (1981) and Rizzi (1982) that apply across clause boundaries, and no formal principle has been proposed, to my knowledge, that predicts or in any way motivates the AH cut-off patterns of (5) and the reverse hierarchy

pattern of Table 1 for gaps and pronouns. There have been isolated attempts in the formal literature to describe the AH cut-off for a particular language in a descriptively adequate way (cf. Cole 1976 for Hebrew). But such descriptions do not explain why the observed universals exist rather than countless others that could just as readily be formalized given current grammatical machinery (e.g. relativization on a DO only, or pronouns high and gaps low, etc). The fact that there is a correlation with patterns of performance and processing complexity is of some theoretical interest, therefore, for the whole question of the origin of grammatical conventions, cf. §0.5.

0.2.2 Patterns in the performance of relative clauses

Some initial performance support for the AH as a complexity ranking was proposed by Keenan & S. Hawkins (1987) on the basis of a repetition experiment conducted on speakers of English, children (11 years) and adults. The prediction was that repetition accuracy would correlate with positions on the hierarchy, subjects being easiest. The data, shown in (9), bear this out (GEN-SU stands for relativization on a genitive within a subject, GEN-DO for relativization on a genitive within a direct object as in (4d)):

(9) Accuracy percentages for English relativizations in a controlled repetition experiment

	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, mostly from English. Wanner & Maratsos (1978) were the first to provide experimental evidence for a measurable processing load within a filler-gap domain, and for the added processing load of DO relatives compared with SU.

Pickering and Shillcock (1992) found significant reaction time differences between the two positions in a self-paced reading experiment, cf. further King & Just 1991, Holmes & O'Regan 1981, Ford 1983, and Hawkins 1999, 2004 for further references and for a metric measuring increasing processing complexity down AH.[1] These experimental results suggest that as the surface domains grow that need to 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 subcategorizor 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 (1999, 2004), which identifies the smallest amount of surface structure containing information sufficient for the unambiguous parsing of a filler-gap dependency.[2]

Some corpus data from Hebrew (Ariel 1999) provide performance support for the grammatical patterns involving gaps versus pronouns presented in §0.2.1. Ariel shows that the Hebrew gap is favored with smaller distances between filler and gap. For example (6a) above, with a minimal distance between filler and gap, is significantly preferred over (6b) with a resumptive pronoun. The pronoun becomes productive when filler-gap domains would be larger, as in (10).

(10) Shoshana hi ha-isha*i* [she-dani siper she-moshe rixel she-nili ohevet ota*i*]

Shoshana is the-woman that-Danny said that-Moshe gossiped that-Nili loves her

This intuition is formalized in Hawkins (2004) by calculating how minimal the total domains can be for Filler-Gap (or Filler-Subcategorizor) processing, for Head-Pronoun co-indexing and for Lexical Argument Structure processing (FGD, HPD and LD respectively):

(10') a Shoshana hi [ha-isha*i* [she-dani siper she-moshe rixel she-nili ohevet 0*i*]]

FGD: ----------

1 2 3 4 5 6 7 8 9 10

LD:ohevet ----------

1 2 3 4 5 6 7 8 9 10

DOMAIN TOTAL = 20

(10') b Shoshana hi [ha-isha*i* [she-dani siper she-moshe rixel she-nili ohevet ota*i*]]

HPD: ----------

1 2 3 4 5 6 7 8 9 10 11

LD:ohevet ----------

1 2 3

DOMAIN TOTAL = 14

In (10'b) the pronoun provides a local argument ota (her) for lexical processing of ohevet (loves), whereas in (10'a) lexical processing needs to access the more distant head ha-isha (woman) in order to assign a direct object to loves. The subject nili is adjacent to ohevet in both cases. More generally, pronoun retention can be hypothesized to reflect the sizes of the domains in which these various relations are processed. The bigger the improvement, the greater will be the preference for the pronoun.[3]

Finally consider some performance data of relevance to AH involving the acquisition of relative clauses, specifically the (second) language acquisition of Swedish by speakers of

languages whose grammars have productive resumptive pronouns in relatives (Persian and Greek) and by speakers whose grammars do not (Spanish and Finnish). Swedish itself has relative clauses not unlike those of English: a relative pronoun co-indexed with the head of the relative is moved to the left of the relative clause, leaving a gap (or subcategorizor) with no resumptive pronoun. Acquisition data quantified by Hylenstam (1984) for the different groups of learners show two clear patterns, cf. Table 2. First, the frequency of resumptive pronouns in Swedish L2 is greater when the L1 has productive pronouns (Persian and Greek), than when it does not. This "transfer effect" is relevant for theories of second language acquisition and confirms its significance among the various factors that shape second language acquisition (cf. the papers in Doughty & Long 2003 and Ramat 2003 for discussion of these factors). Second, what is of significance in the present context is that the general pattern of gaps to pronouns is always the same, regardless of transfer: gaps decline from top to bottom down the AH, while pronouns increase. The absolute quantities for pronouns are higher in the L1s that retain pronouns, but the relative distribution of gaps to pronouns is exactly what we have seen in the grammatical data of Table 1, further confirming the processing basis for gaps in smaller, and pronouns in more complex, environments.

INSERT TABLE 2 HERE

0.3 Minimize Forms

The second principle of efficiency and complexity to be proposed here is (11):

(11) Minimize Forms (MiF)

The human processor prefers to minimize the formal complexity of each

linguistic form F (its phoneme, morpheme, word or phrasal units) and the number of forms with unique conventionalized property assignments, thereby assigning more properties to fewer forms. These minimizations apply in proportion to the ease with which a given property P can be assigned in processing to a given F.

The processing of linguistic forms and of conventionalized property assignments (such as their meanings and syntactic properties) requires effort. Minimizing forms and their property assignments can reduce that effort by fine-tuning it to information that is already active in processing, through accessibility, high frequency, and inferencing strategies of various kinds. According to MiF, minimization is accomplished, first, by reducing the set of formal units in a form or structure, and secondly by reducing the number of forms with unique property assignments.

0.3.1 Form minimization patterns in performance and grammars

Examples abound whose patterning suggests that a reduction in form processing is an advantage, as long as the relevant information can be recovered in processing. Consider the use of pronouns versus full NPs (he/she versus the professor, cf. Ariel's 1990 discussion of high versus low accessibility in discourse correlating with less versus more formal structure respectively), Zipfian (1949) effects (the shorter TV for the high-frequency television), compounds (paper plate for plate made of paper; paper factory for factory that makes paper; cf. Sperber & Wilson's 1995 theory of relevance to, and activation of, real-world knowledge in the processing of minimal structures), co-ordinate deletions (John cooked 0 and Fred ate the pizza), and control structures involving understood subjects of verbs within non-finite subordinate clauses (whose controllers are in a structurally accessible matrix clause position).

Filler-gap dependencies in e.g. relative clauses are also plausibly motivated by (11). Gaps can be identified by reference to their subcategorizor and to the filler with which they are co-indexed. The result is a more minimal structure than resumptive pronoun counterparts, but the advantage of minimalism disappears in complex environments in which processing domains become larger (§0.2.2).

Form reduction is supported further 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).

The minimization principle of (11) adds a second factor to this efficiency logic, beyond the forms themselves, and defined in terms of the properties that are conventionally associated with forms. It is not efficient to have a distinct form F for every possible property P that one might wish to express in everyday communication. 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 then 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}.

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 (Sadock 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. This makes the expression of more frequently used meanings shorter, that of less frequently used meanings longer, which makes communication more efficient overall.

(11) asserts that there is a trade-off between form minimizations 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). This provides a check on how far minimization can go (one cannot minimize everything and assign all

properties through enrichment) and it enables us to make some testable predictions for grammars and performance:

(12) Form Minimization Predictions

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

b The number of unique F:P1 pairings in a language is reduced by grammaticalizing or lexicalizing a given F:P1 in proportion to the frequency and preferred expressiveness of that P1 in performance.

In effect, form minimizations require compensating mechanisms. (12a) asserts that frequency and processing ease regulate reductions in form (their associated properties are more readily inferrable), while frequency and preferred expressiveness regulate the grammaticalization and lexicalization preferences of (12b), which also makes utterances shorter.

0.3.2 Greenberg's Markedness Hierarchies

The effects of these predictions can be seen clearly in Greenberg's (1966) markedness hierarchies such as (13):

(13) Sing > Plur > Dual > Trial/Paucal (for number) [Greenberg 1966, Croft 2003]

Nom/Abs > Acc/Erg > Dat > Other (for case marking) [Primus 1999]

Masc,Fem > Neut (for gender) [Hawkins 1998]

Positive > Comparative > Superlative [Greenberg 1966]

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

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

The other hierarchies had similar frequency correlates. In other words, these hierarchies appear to be **performance frequency rankings** defined on entities within common grammatical and/or semantic domains. The ultimate causes of the frequencies can be quite diverse (real-world frequencies of occurrence, communicative biases in favor of animates rather than inanimates, syntactic and semantic complexity). What is significant for grammars is that these performance rankings are reflected in cross-linguistic patterns that conventionalize morphosyntax and allomorphy in accordance with (12ab).

(15) 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.

(15) follows from (12a). 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 (13). Similarly English singular nouns are zero-marked whereas plurals are formally marked, generally with an -s allomorph.

(16) 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; if it assigns at least one uniquely to B, it does so to A.

(16) follows from (12b). 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.

(17) 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.

(17) also follows from (12b). For example, unique gender-distinctive pronouns can exist for the singular and not for the plural in English (he/she/it vs they), whereas the converse uniqueness is not predicted.

More generally, (16) and (17) lead to a general principle of cross-linguistic morphology:

(18) Morphologization

A morphological distinction will be grammaticalized in proportion to the performance frequency with which it can uniquely identify a given subset of entities {E} in a grammatical and/or semantic domain D.

This principle enables us to make sense of cases of "markedness reversals". For example, in certain nouns in Welsh whose referents are much more frequently plural than singular, like "leaves" and "beans", it is the singular form that is morphologically more complex than the plural, i.e. deilen ("leaf") vs. dail ("leaves"), ffäen ("bean") vs. ffa ("beans"), cf. Haspelmath (2002:244).

0.4 Maximize On-line Processing

The third principle I propose is (19):

(19) 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.

This principle asserts that it is preferable to be able to recognize syntactic and semantic properties efficiently throughout the processing of a sentence and a quantitative metric for measuring this (in terms of OP-to-UP ratios) is proposed in Hawkins (2002, 2004).[4] What is dispreferred is, first, any significant delay or "look ahead" (Marcus 1980) in on-line property assignments, and second any 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 it is also inefficient since initial property assignments are wasted and make no contribution to the ultimate syntactic and semantic representation of the sentence.

0.4.1 Maximize On-line Processing in typological patterns

We see a clear reflex of (19) in a number of patterns across languages that involve asymmetrical ordering preferences between two categories A and B, regardless of the language type. Ordering A before B maximizes on-line processing in these cases, the reverse would involve significant unassignments or misassignments, and MaOP provides a plausible explanation for these conventionalized asymmetries. A sample is given in (20), together with my best estimate of the level of quantitative support for each preference.

(20) a Displaced WH preposed to the left of its (gap-containing) clause [almost exceptionless; §0.4.2, and Hawkins 1999, 2004]

Who*i* [did you say O*i* came to the party]

b Topic to the left of a dependent Predication [exceptionless for some dependencies, highly preferred for others, §0.4.3 and Hawkins 2004]

E.g. Japanese John wa gakusei desu "Speaking of John, he is a student" (Kuno 1973)

c Head Noun (Filler) to the left of its (gap-containing) Relative Clause

E.g. the students*i* [that I teach O*i*]

If a language has basic VO, then N+Relative [exceptions = rare; §0.4.2 and Hawkins 1983, 2004)

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

d Antecedent precedes Anaphor [highly preferred cross-linguistically, §0.4.4]

E.g. **John** washed **himself** (SVO), Washed **John himself** (VSO), **John himself** washed (SOV) = highly preferred over e.g. Washed **himself John** (VOS)

e Wide Scope Quantifier/Operator precedes Narrow Scope Q/O [preferred, §0.4.4]

E.g. **Every student a book** read (SOV languages) ∀∃ preferred

A book every student read (OSV orders in SOV languages) ∃∀ preferred

f Restrictive Relative precedes Appositive Relative (§0.4.4 and Hawkins 2002, 2004)

If N+Relative, then restrictive before appositive relative [exceptionless?]

E.g. Students that major in mathematics, who must work very hard (R+A)

*Students, who must work very hard, that major in mathematics (A+R)

In these asymmetric orders there is an asymmetric dependency of B on A: the gap is dependent on the filler (for gap-filling), the anaphor on its antecedent (for co-indexation), the predication on a topic (for e.g. argument assignment), the narrow scope quantifier on the wide scope quantifier (the number of books read depends on the quantifier in the subject NP in Every student read a book/Many students read a book/Three students read a book), and so on. The assignment of dependent properties to B is more efficient when A precedes, since these properties can be assigned immediately in on-line processing. In the reverse B + A there would be delays in property assignments on-line ("unassignments") or misanalyses ("misassignments"). For example, if the gap were to precede the Wh-word in [you said O*i* came to the party] who*i*, there would be a delay in assigning the subject argument to came; similarly if the predication gakusei desu preceded the topic John wa in Japanese. Let us pursue this idea in more detail.

0.4.2 Fillers First

Gaps are dependent on their fillers for co-indexation and co-reference, and also for recognizing the position to be filled (in conjunction with access to the subcategorizor, 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 (20ac), cf. Hawkins (1999, 2004), J.D. Fodor (1983). When the gap follows the filler, the filler can be fully processed on-line, and the properties that are assigned by reference to the filler can be assigned immediately to the gap on-line, resulting in an efficient distribution of property assignments throughout the sentence. But if the gap precedes, 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 a gap that is activated by late processing of the filler (Antinucci et. al. 1979, Clancy et. al. 1986). Fillers First maximizes on-line property assignments, therefore.

When the filler is a Wh-word in a Wh-question (20a) there is unambiguous cross-linguistic support for Fillers First: almost all languages that move a Wh-word to clause peripheral position move it to the left, not to the right (Hawkins 2004). In relative clauses (20c) there is also clear support, but Fillers First is now in partial conflict with a Minimal Domain preference for noun-final NPs in head-final languages (Hawkins 1994, 2004). Head-initial languages have consistently right-branching relatives (e.g. [V [N S]]), which are motivated both by MiD and by Fillers First. But head-final languages have either left-branching relatives ([[S N] V]), which is good for MiD 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 domains for phrasal processing longer. The

variation here points to the existence of two preferences, whose predictions overlap in one language type but conflict in the other.

The head-final languages that prefer left-branching relatives appear to be the **rigid** ones like Japanese, in which there are more containing head-final phrases (such as V-final VPs) that prefer the head of NP to be final as well (by MiD). **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).

0.4.3 Topics First

A related structure involves topicalized XPs with gaps in a sister S. These generally precede S across languages (Gundel 1988; Primus 1999). The reverse ordering could be optimal for scope marking, but it is either ungrammatical or dispreferred and this provides further evidence for MaOP. 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 preference for Topics First is motivated by the dependence of the gap on the filler for gap identification and filling, as before. In addition the "aboutness" relation between the predication and the topic (Reinhart 1982), coupled with the regular referential independence or givenness of the topic, means that semantic processing of the predication is often incomplete without prior access to the topic, whereas the topic can be processed independently of the predication. For example, Tsao (1978) gives numerous examples from Mandarin Chinese of a 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. These examples include:

Argument assignment to, and disambiguation of, the subcategorizor in the predication:

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

Various **argument enrichments** whereby the topic provides a **possessor** (22), **class** (23), **set** (24) or **restrictive adjunct** (25) relative to which an argument in the predication is interpreted:

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

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

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

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

Various **predicate enrichments** whereby the topic provides a **location** (26), **time** (27), or **cause** (28) adjunct, or a domain for **superlative** (28) interpretation relative to which the predication is interpreted.

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

(27) **Dzwo-tyan** (**a**), Jang San lai kan wo. (predicate enrichment:

Yesterday (Topic Part), Jang San came (to) see me. time

(28) **Weile jei-ge haidz**, wo bu jr chr-le dwoshau ku. (predicate enrichment:

For (/on account of) this-Classif child, I have endured much hardship. cause)

(29) **Yu (a)**, wei-yu syandzai dzwei gwei. (predicate enrichment:

Fish (Topic Part), tuna is now the most expensive. superlative domain)

If predication and topic were reversed in these examples, 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, i.e. there would be online unassignments and misassignments. In e.g. (22) it would be unclear whose mind was intended, in (23) 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 (29) the expensiveness of tuna must be interpreted relative to fish, not say food in general, and unless this restriction is contextually given it cannot be assigned online when fish follows.

These asymmetries predict a topic + predication ordering preference, thereby avoiding temporary unassignments or property misassignments on-line. Across languages, argument enrichments and predicate enrichments (i.e. with fully asymmetric dependencies) appear to be entirely topic + 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 theta-role assignment) are preferably topic + predication (Hawkins 2004).

0.4.4 Other linear precedence asymmetries

Further ordering asymmetries that are plausibly motivated by MaOP include the preference for antecedents before their anaphors (dependent on the former for co-indexing and co-reference (20d)), and wide scope before narrow scope operators and quantifiers (20e). Positioning the wide scope item first permits immediate assignment of the appropriate interpretation to the narrow scope item, by reference to the already processed wide scope item, and avoids un/misassignments on-line. Compare the different interpretations of the indefinite singular a book in All the students read a book/Some students read a book/Three students read a book). When a book precedes (A book all the students read, etc) there is no higher scope element in working memory relative to which a narrow scope interpretation can be assigned, and the preferred interpretation shifts to wide scope.

Also relevant here is the preference for restrictive before appositive relatives exemplified by (30) in English (cf. (20f)):

(30) 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 (30b) there would always be a semantic garden path. The appositive relative would first be predicated of all students, and would then be revised to a predication about that subset only of which the restrictive relative was true, once the latter was encountered and processed. The ordering of (30a) avoids the regular garden path by placing together all the items that determine the reference set of which the appositive clause is predicated, positioning them before the appositive claim in surface syntax. R+A appears to be widespread in head-initial languages. For head-final languages, cf. Hawkins (2004:241) and Lehmann (1984:277-80).

Notice finally that in contrast to the asymmetrical dependencies of (20), dependencies between a verb and e.g. an NP direct object are symmetrical. NP depends on V for case- and theta-role assignment and also for mother node recognition (VP) and attachment (Hawkins 1994), while V depends on NP for selection of the intended syntactic and semantic co-occurrence frame (e.g. transitive vs. intransitive run [John ran/John ran the race]), and for the intended semantics of V from among ambiguous or polysemous alternatives (ran the race/the water/the advertizement/his hand through his hair, Keenan 1979). These symmetrical dependencies are matched by symmetrical ordering patterns across languages (A+B/B+A), e.g. VO & OV. Asymmetrical orderings appear to involve strong asymmetries in dependency, therefore (as defined here in processing terms, cf. (2) above), whereas symmetrical dependencies result in symmetrical orderings (Hawkins 2004).

0.5 Conclusions

I conclude that typological patterns can be profitably described, predicted, and to a significant extent explained in terms of principles of efficiency and complexity in processing. More generally I have proposed a Performance-Grammar Correspondence Hypothesis (1) whereby preferences in performance (in languages with variation) are matched by conventionalized structures in grammars. Three general principles have been proposed: Minimize Domains (§0.2); Minimize Forms (§0.3); and Maximize On-line Processing (§0.4). 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. They also explain numerous typological patterns of the kind summarized here, many of which are either not predicted by grammar-only principles or that provide frequent exceptions to generative parameters (cf. Newmeyer 2005, Hawkins 2004).

Conversely these patterns become relevant for theories of processing (and acquisition, cf. §0.2.2), since grammars are hereby claimed to be conventionalizations of the same processing mechanisms that psychologists find evidence for in experimental and corpus data. Grammatical patterns can suggest principles for testing in relevant languages (e.g. the Accessibility Hierarchy in §0.2 led to predictions for processing and acquisition), and they can provide a check on psycholinguistic hypotheses (many of which are still too Eurocentric and based on an insufficient sample of the world's languages) and can lead to improved processing theories, cf. Hawkins 2004 and Yamashita & Chang 2001.

It remains to be seen how much of classic typology (and of core syntax and syntactic variation within generative grammar) can be explained in terms of the PGCH (1). I believe the examples we have seen are just the tip of a large iceberg. And if these performance-grammar correspondences are valid, then any explanation that accounts for grammars only, as in the Chomskyan philosophy of grammar (Chomsky 1965, 1986, Hoekstra & Kooij 1988), will be missing significant generalizations. The alternative proposed here 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 underlies these ease of use and efficiency regularities is most plausibly innate.

It will ultimately be necessary to answer some general questions that are raised by this 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 like those proposed here and that different constraint rankings and outputs can become conventionalized through a process of diachronic adaptation. There are also psycholinguistic issues that are raised by these performance preferences. How exactly can these efficiencies 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 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 results in the typological patterns that we have seen here. And studying these patterns from the perspective of the PGCH (1) results in an interdisciplinary research program that we can call "Processing Typology", in whose pursuit I invite all interested parties to join me.

References

Antinucci, F., Duranti, A. and Gebert, L. (1979). 'Relative Clause Structure, Relative Clause Perception and the Change from SOV to SVO', Cognition 7: 145-176.

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

Ariel, M. (1999). 'Cognitive Universals and Linguistic Conventions: The Case of Resumptive Pronouns', Studies in Language 23:217-269.

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

Chomsky, N. (1965). Aspects of the Theory of Syntax. Cambridge, Mass.: 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. and Zoh, M. (1986). 'Processing Strategies in the Acquisition of Relative Clauses', Cognition 14: 225-262.

Cole, P. (1976). 'An Apparent Asymmetry in the Formation of Relative Clauses in Modern Hebrew', in P. Cole (ed.), Studies in Modern Hebrew Syntax and Semantics. Amsterdam: North Holland.

Comrie, B. (1989). Language Universals and Linguistic Typology, 2nd edn. Chicago: University of Chicago Press.

Croft, W. (2003). Typology and Universals, 2nd edn. Cambridge, Cambridge University Press.

Doughty, C.J. and Long, M.H. (eds.). (2003). The Handbook of Second Language Acquisition. Oxford: Blackwell Publishing.

Fodor, J.D. (1983). 'Phrase Structure Parsing and the Island Constraints', Linguistics and Philosophy 6: 163-223.

Ford, M. (1983). 'A Method of Obtaining Measures of Local Parsing Complexity throughout Sentences', Journal of Verbal Learning and Verbal Behavior 22: 203-218.

Gell-Mann, M. (1992). 'Complexity and Complex Adaptive Systems', in J.A. Hawkins and M. Gell-Mann (eds), The Evolution of Human Languages. Redwood City, CA: Addison-Wesley.

Gibson, E. (1998). 'Linguistic Complexity: Locality of Syntactic Dependencies', Cognition 68: 1-76.

Givón, T. (1979). On Understanding Grammar. New York: Academic Press.

Greenberg, J.H. (1966). Language Universals, with Special Reference to Feature Hierarchies. The Hague: Mouton.

Greenberg, J.H. (1995). 'The Diachronic Typological Approach to Language', in M. Shibatani and T. Bynon (eds), Approaches to Language Typology. Oxford: Clarendon Press.

Grice, H.P. (1975). 'Logic and Conversation', in P. Cole and J. Morgan (eds.), Speech Acts. New York: Academic Press, 41-58.

Gundel, J.K. (1988). 'Universals of Topic-Comment Structure', in M. Hammond. E. Moravcsik and J.R. Wirth (eds.), Studies in Syntactic Typology. Amsterdam: John Benjamins, 209-239.

Haiman, J. (1983). 'Iconic and Economic Motivation', Language 59: 781-819.

Haspelmath, M. (1999). 'Optimality and Diachronic Adaptation', Zeitschrift für Sprachwissenschaft 18: 180-205.

Haspelmath, M. (2002). Morphology. London: Arnold.

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. (1998). 'Morphological Hierarchies in the History of Germanic', International Journal of Germanic Linguistics and Semiotic Analysis 3:197-217.

Hawkins, J.A. (1999). 'Processing Complexity and Filler-Gap Dependencies', Language 75:244-285.

Hawkins, J.A. (2002). 'Symmetries and Asymmetries: Their Grammar, Typology and Parsing', Theoretical Linguistics 28:95-149.

Hawkins, J.A. (2004). Efficiency and Complexity in Grammars. Oxford: Oxford University Press.

Hoekstra, T. and Kooij, J.G. (1988). 'The Innateness Hypothesis', in J.A. Hawkins (ed.), Explaining Language Universals. Oxford: Basil Blackwell.

Holmes, V.M. and O'Regan, J.K. (1981). 'Eye Fixation Patterns during the Reading of Relative Clause Sentences', Journal of Verbal Learning and Verbal Behavior 20:417-430.

Hyltenstam, K. (1984). 'The Use of Typological Markedness Conditions as Predictors in Second Language Acquisition: The Case of Pronominal Copies in Relative Clauses', in R.W. Andersen (ed.), Second Languages: A Cross-linguistic Perspective. Rowley, Mass.: Newbury House, 39-58.

Jaeger, T.F. and Wasow, T. (2005). 'Production-complexity Driven Variation: Relativizer Omission in Non-subject-extracted Relative Clauses'. Paper presented at the 18th Annual CUNY Sentence Processing Conference, Tuscon, AZ.

Keenan, E.L. (1979). 'On Surface Form and Logical Form', Studies in the Linguistic Sciences 8: 163-203.

Keenan, E.L. and Comrie, B. (1977). 'Noun Phrase Accessibility and Universal Grammar', Linguistic Inquiry 8:63-99.

Keenan, E.L. & Hawkins, S. (1987). 'The Psychological Validity of the Accessibility Hierarchy", in E.L. Keenan, Universal Grammar: 15 Essays. London: Croom Helm.

King, J. & Just, M.A. (1991). 'Individual Differences in Syntactic Processing: The Role of Working Memory', Journal of Memory and Language 30: 580-602.

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, Mass.: MIT Press.

Lehmann, C. (1984). Der Relativsatz. Tübingen: Gunter Narr Verlag.

Levinson, S.C. (2000). Presumptive Meanings: The Theory of Generalized Conversational Implicature. Cambridge, Mass.: MIT Press.

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

MacDonald, M.C., Pearlmutter, N.J. and Seidenberg, M.S. (1994). 'The Lexical Nature of Syntactic Ambiguity Resolution", Psychological Review 101.4: 676-703.

Marcus, M.P. (1980). A Theory of Syntactic Recognition for Natural Language. Cambridge, Mass.: MIT Press.

Matthews, S. and 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.

Newmeyer, F.J. (2005). Possible Languages and Probable Languages. Oxford: Oxford University Press.

Pickering, M. and Shillcock, R. (1992). 'Processing Subject Extractions', in H. Goodluck and M. Rochemont (eds.), Island Constraints: Theory, Acquisition and Processing. Dordrecht: Kluwer, 295-320.

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

Race, D.S. and MacDonald, M.C. (2003). 'The Use of "that" in the Production and Comprehension of Object Relative Clauses". Paper presented at the 26th Annual Meeting of the Cognitive Science Society.

Ramat, A.G. (ed.). (2003). Typology and Second Language Acquisition. Berlin: de Gruyter.

Reinhart, T. (1982). 'Pragmatics and Linguistics: An Analysis of Sentence Topics', Philosophica 27:53-94.

Rizzi, L. (1982). 'Violations of the WH-island Constraint and the Subjacency Condition', in L. Rizzi (ed.), Issues in Italian Syntax. Dordrecht: Foris, 49-76.

Ross, J.R. (1966). 'Constraints on Variables in Syntax', Ph.D. dissertation, MIT.

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

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

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

Tsao, F-F. (1978). A Functional Study of Topic in Chinese. Taipei: Student Book Company.

Wanner, E. and Maratsos, M. (1978). 'An ATN Approach to Comprehension', in M. Halle, J. Bresnan and G.A. Miller (eds.), Linguistic Theory and Psychological Reality. Cambridge, Mass.: MIT Press, 119-161.

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

Yamashita, H. and 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.

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

Table 1 Languages combining [-Case] gaps with [+Case] pronouns (Keenan & Comrie 1977)

	SU	DO	IO/OBL	GEN
Aoban	gap	pro	pro	pro
Arabic	gap	pro	pro	pro
Gilbertese	gap	pro	pro	pro
Kera	gap	pro	pro	pro
Chinese (Peking)	gap	gap/pro	pro	pro
Genoese	gap	gap/pro	pro	pro
Hebrew	gap	gap/pro	pro	pro
Persian	gap	gap/pro	pro	pro
Tongan	gap	gap/pro	pro	pro
Fulani	gap	gap	pro	pro

Greek	gap	gap	pro	pro
Welsh	gap	gap	pro	pro
Zurich German	gap	gap	pro	pro
Toba Batak	gap	*	pro	pro
Hausa	gap	gap	gap/pro	pro
Shona	gap	gap	gap/pro	pro
Minang-Kabau	gap	*	*/pro	pro
Korean	gap	gap	gap	pro
Roviana	gap	gap	gap	pro
Turkish	gap	gap	gap	pro
Yoruba	gap	gap	0	pro
Malay	gap	gap	RP	pro
Javanese	gap	*	*	pro
Japanese	gap	gap	gap	gap/pro

Gaps =	24 [100%]	17 [65%]	6 [25%]	1 [4%]
Pros =	0 [0%]	9 [35%]	18 [75%]	24 [96%]

Key: gap = [-Case] strategy

pro = copy pronoun retained (as a subinstance of [+Case])

* = obligatory passivization to a higher position prior to relativization

0 = position does not exist as such

RP = relative pronoun plus gap (as a subinstance of [+Case])

[-Case] gap languages may employ a general subordination marker within the relative clause, no subordination marking, a participial verb form, or a fronted case-invariant relative pronoun. For Tongan, an ergative language, the top two positions of AH are Absolutive and Ergative respectively, not SU and DO, cf. Primus (1999).

Table 2 Gaps and Pronouns in Swedish Second Language Acquisition (Hyltenstam 1984)

Gaps	SU	DO	IO/OBL	GEN
PersianLSwedish	100%	42%	25%	8%
GreekLSwedish	100%	58%	42%	8%
SpanishLSwedish	100%	83%	62%	8%
FinnishLSwedish	100%	100%	100%	33%
Pronouns				
PersianLSwedish	0%	58%	75%	92%
GreekLSwedish	0%	42%	58%	92%
SpanishLSwedish	0%	17%	38%	92%
FinnishLSwedish	0%	0%	0%	67%

Footnotes

1. Gibson's (1998) "locality" principle makes many similar predictions to those of MiD and the wealth of experimental support that he summarizes there carries over to the MiD.

2. A Filler-Gap Domain (FGD) is defined as follows in Hawkins (1999): 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 subcategorizor and includes, or is extended to include, the gap's dependent and disambiguating arguments (if any); for nonsub-categorized gaps the path connects the filler to the gap site; all constituency relations and coocurrence requirements holding between these nodes belong to the description of the FGD.

3. There are other processing factors that impact preferences for relative clause variants, beyond minimal domains of the kind defined here. E.g. the overall size and complexity of a relative clause leads to a preference for the explicit relative pronoun in English (versus zero),

even when additional material in the relative is in post-verbal (or post-gap) position and falls outside the filler-gap and lexical domains of Hawkins (2004)

, cf. Race & MacDonald (2003), Jaeger & Wasow (2005). There are also more resumptive pronouns in adjunct rather than argument positions in Hebrew, and in non-restrictive versus restrictive relatives (Ariel 1999). Domain minimization is just one pattern predictor, therefore, and it remains to investigate whether grammars have responded to the other patterns as well. Some factors, such as overall terminal length of the relative, will be harder to grammaticalize, for reasons discussed in Hawkins (1994:19-24).

4. Notice that Maximize On-line Processing is formulated in terms of parsing and the hearer, since the speaker does not make structural missassignments on-line and can enrich unassignments based, inter alia, on knowledge of what is to be produced later.