Function, Selection and Innateness

The Emergence of Language Universals

Simon Kirby

A thesis submitted in fulfilment of the requirements

for the degree of Doctor of Philosophy

to the

University of Edinburgh

1996

Abstract

A central topic for linguistic theory is the degree to which the communicative function of language influences its form. In particular many so-called functional explanations argue that cross-linguistic constraints can be explained with reference to pressures imposed by processing. In apparent opposition to this is the innatist stance which claims that universals are properties imposed by an autonomous language module. This thesis approaches the issues raised by this conflict by examining the nature of the link between processing and universals. The starting point for the work, then, is not the discovery of new universals nor new explanations, but the question "exactly how do processing theories that have been proposed give rise to the universals that they claim to explain?" Careful investigation of this problem proves to be fruitful in highlighting the roles of innateness and function in explaining universals.

The methodology chosen involves computational simulations of language as a complex adaptive system, in which language universals appear as emergent properties of the dynamics of the system and the influence of processing on use. This influence is characterised as a differential selection of competing variant forms. The simulation approach is first used to demonstrate the plausibility of a recent parsing explanation for word order universals. An extension of the model to deal with hierarchical universals relating to relative clauses leads to the conclusion that current explanations of hierarchies in general are incomplete. Instead, it is argued that implicational hierarchies are the result of competing processing pressures, in particular between morphological and parsing complexity.

Further examination of relative clause processing and universals leads to an apparent flaw in the approach put forward. It is noted that not all processing pressures appear to show up as universals, challenging the explanatory adequacy of the functional explanations. Instead, it is shown that a complete characterisation of language as an adaptive system requires there to be an innate, autonomous syntactic component to language. This leads to the conclusion that universals arise from the interaction of processing constraints and constraints imposed on the adaptive process by an innate language acquisition device. Moreover, the possibility of processing directly influencing this innate faculty without violating its autonomy is investigated with reference to recent work on the biological evolution of language.

This thesis therefore espouses a perspective on the explanation of language universals in which processing complexity and autonomous syntactic constraints have crucial and complementary roles.

Declaration

This thesis has been composed by myself and it has not been submitted in any previous application for a degree. The work reported within was executed by myself, unless otherwise stated.

April 1996

Acknowledgements

Before I started writing up, I was under the impression that it would be an extremely painful activity. I'm glad to say that, surprisingly, this wasn't the case. However, I feel sure that this was largely due to the assistance and generosity of those around me (some of whom would probably disagree with me on how painless the whole thing was!).

Firstly, thanks goes to my supervisor, Jim Hurford, whose patience with my continual requests and readiness to engage in impromptu supervisions at random times, have kept my interest and excitement in the subject alive. Thanks are due to the linguists who have taken time to comment on my work, send me papers, or otherwise assist and give encouragement. In no particular order, these include: Jack Hawkins, Matthew Dryer, Louise Kelly, Diane Nelson, Daniel Wedgwood, Caroline Heycock, Jean Aitchison, Steve McGill, Ronnie Cann, Margaret Winters, Geoff Nathan, Dik Bakker, Frans Plank, Bob Ladd, Maggie Tallerman, Anna Siewierska, Steven Pinker and several anonymous reviewers. Not all of these people will agree with what I've said, of course, and almost none of them are responsible for any mistakes.

The research for and preparation of this thesis was made possible in part because of the quality of the software that was used, almost all of which was produced for free in the spirit of GNU and Linux. The many people who work on these projects are amassing a great debt of gratitude.

A large part of what made the last three years enjoyable is the nature of the Department of Linguistics at Edinburgh. Thanks to those that keep it from falling apart around us, especially Ethel Jack, Irene McLeod and Cedric MacMartin (again, for incredible patience). Central to the department is the Common Room, and all those who have made it more than just a place to eat noodles, particularly: Dave (rowing), Dave (hair), Dan, Catriona, Louise, Miriam, Anna, Etsuko and Julie.

If all I had done over the past three years was work on this thesis, I would clearly have lost the plot long ago, so a heartfelt thankyou goes to the bands I've played in during this period. There's nothing quite so different from writing a thesis than playing funk in a pub. So, to everyone in Kettlefish, Big Sur and the Ugly Groove Movement (your names in full would cost too much): cheers, and apologies for echo-guitar, 7/8, and memory loss respectively. Apart from all these people, there are several others without whom I would have struggled to survive. A particularly large round of drinks to: Helen, Anna, Brian, Gav, Tuna, Ann, Matt, Paul, Al, and Ian (I owe you 4,000 cups of coffee). And for unquestioning support: thanks and much love to my parents and Diane.

Finally, to John, one last drink for all the ways in which you helped me in the past; I only wish you could have stayed.

Abbreviations

AH .... accessibility hierarchy
ASP .... aspect
Adj .... adjective
BDT .... branching direction theory
C′ .... complementiser-bar
CP .... complementiser phrase
CRD .... constituent recognition domain
Comp .... complementiser
D .... determiner
D′ .... determiner-bar
DO .... direct object
DP .... determiner phrase
Det .... determiner
EIC .... early immediate constituent (recognition)
GB .... government and binding theory
Gen .... genitive
HPSG .... head-driven phrase structure grammar
I .... indirect object relative
I .... inflection
I′ .... inflection-bar
IC .... immediate constituent
IO .... indirect object

IP . . . . . inflection phrase
LAD . . . . . language acquisition device
LF . . . . . logical form
MNCC . . . . . mother node constructing category
MSG . . . . . male-singular
MUT . . . . . mutation
Mod . . . . . modifier
N . . . . . noun
NP . . . . . noun phrase
O . . . . . object relative
OBL . . . . . oblique
OV . . . . . object-verb
P . . . . . preposition/postposition
PF . . . . . phonetic form
PFLP . . . . . phonetic form licensing principle
PLD . . . . . primary linguistic data
PP . . . . . adpositional phrase
Po . . . . . postposition(al)
Postp . . . . . postposition(al)
Pr . . . . . preposition(al)
PrNMH . . . . . prepositional noun-modifier hierarchy
Pref . . . . . prefix
Prep . . . . . preposition(al)
Pron . . . . . pronoun
RC . . . . . relative clause
Rel . . . . . relative
S . . . . . sentence
S . . . . . subject relative
S′ . . . . . sentence-bar
SG . . . . . singular

SOV .... subject-object-verb
SVO .... subject-verb-object
Spec .... specifier
Suff .... suffix
UG .... universal grammar
V .... verb
VO .... verb-object
VP .... verb phrase
VSO .... verb-subject-object

Contents

List of Figures

Chapter 1

A puzzle of fit

A striking feature of the natural world is the existence of organisms whose occurrence is improbable simply by virtue of their complexity.[1] Matter seems to arrange itself into highly organised bundles whenever life intervenes. The examples of this improbable order extend to the artifacts of life as well as living things themselves: for example, the buildings, roads and pavements that make up towns and, more abstractly, the cultural patterns that give rise to these artifacts. All of these things are improbable in the sense that they inhabit a small, organised area in the space of logical possibility.

This thesis looks at another phenomenon in the natural world: human language. The range of variation among languages is constrained in various interesting ways expressed by descriptive statements of "language universals". These universals map the boundaries of a small area in the space of logically possible languages within which the actual languages of the world are found. In other words, languages do not randomly vary from instance to instance but rather embody a kind of pattern and ordered complexity similar to that found in life and its other artifacts.

The origin of this order is in itself interesting, but I shall be exploring a particular aspect of these constraints on variation that they share with others in the natural world. This aspect can be termed "fit" or "the appearance of design". Trees appear to

[1] No definition of this type of complexity is given here. Algorithmic complexity is not a good definition, since some organised, complex distributions (e.g. fractal sets) can be defined quite simply. See, for example, Gell-Mann 1992 for some discussion.

be designed for the purpose of surviving in the world and producing more trees — looking deeper, we can say they appear to be designed for converting carbon dioxide and sunlight into more tree, and so on. Buildings appear to be designed to efficiently contain people and their possessions without harm from the weather (in fact, we know they are designed for this purpose). As Cziko (1995) (from whom this chapter title is borrowed) points out, this "fit" of form to function pervades the world of living organisms and their products.

As we shall see, this appearance of design is also a striking feature of language universals. Many attempts at explaining universals have pointed out the fit of these constraints of variation to the functions of language. Although these observations are important and insightful, I believe they leave the real mystery unsolved. Rather than explaining the origin of universals, this fit is itself a puzzle. Where does it come from, and what mechanisms can explain how it arises? A careful study of this question casts light on many issues in modern linguistics and reflects back on the various views of what makes a "possible human language".

1.1 Constraints on variation

Before turning to possible explanations for universals, it is worth looking at the various forms these constraints on linguistic variation take. The formulation of a language universal involves two distinct steps:

Typology This is a taxonomy which categorises languages along some dimension on the basis of an identifiable property of the language. For the purposes of formulating a universal, orthogonal typologies may be considered, leading to a particular language being placed in a multi-dimensional space.

Constraints The actual universal is stated as a constraint on possible language types, defining a sub-space within the space defined by the typology.

1.1.1 Forms of constraints

The constraints may take various forms, which can be usefully categorised on two dimensions (notice that the broad distinctions here are recognised by other authors (e.g. Greenberg 1963; Comrie 1981; Hawkins 1988; Croft 1990) although the precise formulation is not identical). Firstly, the constraints may be **absolute**[2] or **statistical**. In other words, they can differ in the degree to which we may expect exceptions. This might immediately seem problematic since how can we state a constraint on possible human languages that may be violated? However, it is important to realise that a statistically significant skewing of the distribution of languages is as worthy of explanation as an absolute one.[3]

Secondly, the format of the constraint can typically be characterised as **parametric** or **hierarchical**. This difference is related to the logical relationships between typological dimensions:

Parametric universals These describe a co-occurrence relation between different types, so that when one type occurs, so does the other *and vice versa*. They are expressed logically as:[4]

$$\forall L[(P_1(L) \leftrightarrow P_2(L))\&(P_2(L) \leftrightarrow P_3(L))\&\ldots\&(P_{n-1}(L) \leftrightarrow P_n(L))]$$

where P_i is some property of a language L that differentiates between a type T_i and T_i', where a prime here indicates an opposite type.[5]

[2] The term *absolute universal* is sometimes used for substantive or formal universals that simply constrain languages to all have a certain property.

[3] This leads to the problem of identifying statistical significance (as will be discussed in chapter 2), but this problem is equally present for absolute universals. For example, imagine a typology categorises languages into 3 types: A, B and C. Let us say in a typologist's sample that 99% of languages are type A, 1% are type B and none are type C. From an absolute stance, we would conclude that human languages can be A or B but never C. However, what if type C was simply missing from the sample but observable elsewhere? If this were the case, then A, B and C should be given the same status in absolute terms. A statistical approach, on the other hand would enable us to say that A was significantly more common than B or C.

[4] For convenience we can simply abstract away from L in the expression of these universals in other places in this thesis.

[5] This formulation relies on a binary typology. However, other typologies can be easily reduced to this case.

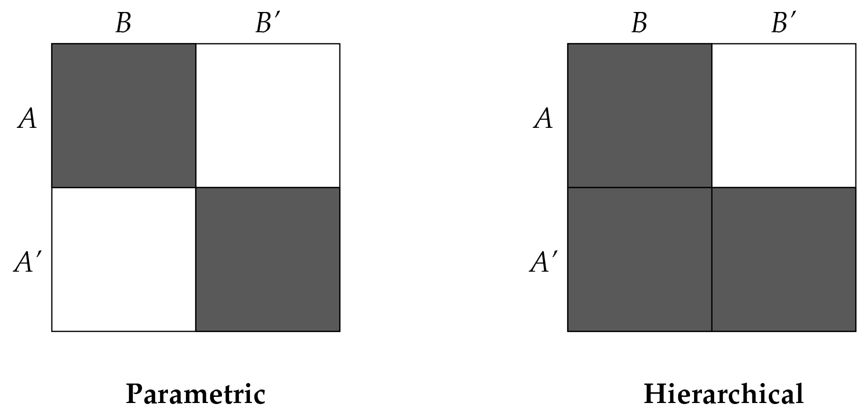


Figure 1.1. Constraints on variation in parametric and hierarchical universals involving types A and B.

Hierarchical universals These also describe co-occurrence relations, but crucially they are asymmetric across types. The logical expression is as:

$$\forall L[(P_1(L) \rightarrow P_2(L))\&(P_2(L) \rightarrow P_3(L))\& \ldots \&(P_{n-1}(L) \rightarrow P_n(L))]$$

The simplest hierarchical universal involving two type dimensions is traditionally termed an *implicational universal*. These may also be written using the symbol $\supset$ instead of $\rightarrow$.

The difference between hierarchical and parametric can be seen diagrammatically in figure 1.1 for the simplest case of two binary types. In general, parametric universals constrain attested languages to 2 in 2^n possibilities, and hierarchical universals constrain to $n+1$ in 2^n, so even for a small number of types, these universals are highly predictive.

Both types of universal can be found in Greenberg's (1963) pioneering work. For example (using Greenberg's numbering):

(27) If a language is exclusively suffixing, it is postpositional. If it is exclusively prefixing, it is prepositional.

(3) Languages with dominant VSO order are always prepositional.

The first two universals together parametrically relate affixation with adpositional order for languages whose affixes exclusively pattern one way or the other. This can be written $Pref \leftrightarrow Prep$, where $Pref' \equiv Suff$ and $Prep' \equiv Postp$. The universal (3) is different in that it does not rule out a prepositional language that is not VSO, such as English, so it would be expressed as $VSO \rightarrow Prep$.

1.1.2 Hierarchies

The second type of universal is of special interest to linguists as it defines an asymmetrical hierarchy of types. These are often written using the $>$ operator to express relative height on the hierarchy. A universal such as:

$$(A \rightarrow B)\&(B \rightarrow C)$$

would be written:

$$C > B > A$$

Languages can be defined by their position on such a hierarchy, since any language with a property corresponding to a type low on the hierarchy will also have the properties of the types higher on the hierarchy. The Greenberg universal (3) above could be expressed as a hierarchy $Prep > VSO$, and English could be placed half way up this hierarchy as having $Prep$ but not VSO. This is not usually done for such simple implicational universals, however. Instead, the hierarchy is reserved for "chained implications" or multi-typed hierarchical universals in our terms.

The paradigm case of an implicational hierarchy is the Keenan & Comrie (1977) Accessibility Hierarchy, which is based on a typology of languages relating to their ability to relativise various grammatical functions within a subordinate clause (with a "basic" strategy — see chapter 3 for more detail). A portion of this hierarchy is:

$$DO > IO > OBL$$

where DO=Direct Object, IO=Indirect Object, and OBL=Oblique. This corresponds to the universal:

$$(OBL \rightarrow IO)\&(IO \rightarrow DO)$$

In other words, if a language allows oblique relatives, then it allows indirect object relatives, and if it allows indirect object relatives, it allows direct object relatives. Notice, that these are binary types which refer to relativisation or lack of relativisation for each grammatical function. The type IO has its counterpart IO' which is assigned to a language which *cannot* relativise indirect objects (such as Yoruba). This hierarchy constrains human languages according to the following table (where each row is a possible language and + means that a language is of a particular type):

DO	IO	OBL	Example
–	–	–	Batak
+	–	–	Hausa
+	+	–	Catalan
+	+	+	English

Although not required by the logical structure of a universal, a typically unspoken requirement of a hierarchy such as this is that there is an example language for each position on the hierarchy. In fact Keenan & Comrie (1977) make this explicit in their original formulation of this particular universal: "Strategies that apply at one point on the [Accessibility Hierarchy] may in principle cease to apply at any lower point."

Each implicational statement has a logical equivalent related to it by *modus tollens*. The implication $P \rightarrow Q$ is identical, truth conditionally, to $\neg Q \rightarrow \neg P$. In terms of binary types, this means that if $A \rightarrow B$ is a universal, then so is $B' \rightarrow A'$. I will refer to this as the **contrapositive universal**. The hierarchy above thus has a contrapositive equivalent:

$$OBL' > IO' > DO'$$

In other words, if a language cannot relativise direct objects, then it cannot relativise indirect objects, and if a language cannot relativise indirect objects, it cannot relativise direct objects. In chapter 3, the choice of a hierarchy or its contrapositive "twin" will be shown to reflect on its explanation. The contrapositive table of possible languages is simply a mirror image of the one above:

OBL′	IO′	DO′	Example
–	–	–	English
+	–	–	Catalan
+	+	–	Hausa
+	+	+	Batak

1.2 The evidence of fit

I have said that language universals show the "appearance of design" in that there is a fit of form to function. The search for this fit underlies an approach to the explanation of universals that is usually referred to as the *functional approach*. This term appears to be used mainly to set up an opposition between linguists interested in language function and those following the *generative* or *formal* approach (to which we will turn shortly). However functionalism does not consist of a single, coherent research program; rather it characterises any attempt to explain universals in terms of language use.

1.2.1 Types of functional explanation

Various authors, in reviewing explanations for language universals, have pointed out the different aspects of language use that have been called upon in functional explanation (see, e.g. Comrie 1981, 26–29; Hawkins 1988, 8–18; Hurford 1990, 94–96; Croft 1990, 252–256; Hall 1992, 27–32, and references therein). For example, Comrie (1981:28) notes that "the existence of first or second person reflexive forms in a language implies the existence of third person reflexive forms". He appeals to **pragmatics** to explain this constraint. In an English utterance, different instances of *I* always refer to the same entity. Similarly, almost all instances of *we* or *you* will be coreferential. On the other hand, third person pronouns are regularly non coreferential in an utterance. Comrie suggests that the reflexive/non-reflexive distinction is therefore more important functionally for third person referents than first or second person referents.

Another type of explanation appeals to **iconicity**, or the isomorphism of sign and signified. Greenberg's (1963) universal (28) states "if both the derivation and inflection follow the root, or they both precede the root, the derivation is always between the root and the inflection". Bybee's (1985) explanation for this is that the formal closeness of an affix to its stem iconically reflects its conceptual closeness — the degree to which the semantics of the affix affects solely the meaning of the word. In Croft's (1990) words, "derivational morphology alters the lexical meaning of the root, sometimes drastically, whereas inflectional morphology only adds semantic properties or embeds the concept denoted by the root into the larger linguistic context"(p.176).

A third type of explanation appeals to the structure of **discourse**. An interesting and complex example is DuBois'(1987) explanation of the tendency for languages' case systems to pattern as *ergative* or *nominative-accusative*. Briefly, the nominative-accusative pattern, which reserves special marking for the object of a transitive as opposed to the subject of transitives and intransitives, represents an iconic patterning of agents versus non-agents in language. The ergative system, on the other hand matches a preferred argument structure in discourse. DuBois shows with text counts that most clauses in discourse involve only one or zero nominal arguments. This is because transitive subjects are usually "given" topics and therefore pronominal. This means that full-NPs are most often subjects of intransitives or objects of transitives, hence the special marking reserved for subjects of transitives in ergative case systems. DuBois goes on to extend his analysis to split-ergative patterns, but a full treatment of his approach would be beyond the purposes of this review.

Finally, **processing** has often been appealed to in the explanation of universals. Cutler *et al.* (1985) aim to explain the cross-linguistic preference for suffixes (as opposed to prefixes) in terms of the way in which language is processed by hearers in real time. The crucial feature of this processing is that it is constrained by the left-to-right, serial nature of speech. The start of a word is clearly received by the processor before the end, and the assumption is that work starts on processing input as soon as it arrives. Simplifying the situation somewhat, Cutler *et al.* point out that early lexical access is preferred by hearers so the placing of salient information early in the word aids processing. If lexical access is stem-based — as they argue from experimental evidence — then the tendency for languages to be suffixal matches the preference of the processor.

1.2.2 Aspects of function

The brief review above highlights the main feature functional explanations have in common: universals are "explained" by demonstrating that their content matches some feature of language use. Typically, some difference between pairs of linguistic objects matches a similar difference in the use of those objects (where objects here is taken to mean anything that corresponds to a *type*). So, differences between reflexives of second and third person correspond to differences in the use of those reflexives in utterances. Differences in the position of derivational and inflectional affixes correspond to differences in the use of those affixes to signal changes in meaning. The differential marking of transitive subjects in ergative languages corresponds to their special role in discourse. The difference in the distribution of suffixes and prefixes cross-linguistically mirrors the left-to-right processing of words. In this way, all these explanations appeal to the fit of universals to function.

However, we have so far been rather vague about what constitutes "function". The explanations above rely on features of language use, but there appear to be rather different aspects of use that are typically emphasised. For example, Hyman (1984) makes a distinction between two types of function:

> "Unfortunately, there is disagreement on the meaning of 'functional' as applied in this context. While everyone would agree that explanations in terms of communication and the nature of discourse are functional, . . . explanations in terms of cognition, the nature of the brain, etc., are considered functional by some but not other linguists. The distinction appears to be that cognitive or psycholinguistic explanations involve formal operations that the human mind can vs. cannot accommodate or 'likes' vs. 'does not like', etc., while pragmatic or sociolinguistic explanations involve (formal?) operations that a human society or individual within a

> society can vs. cannot accommodate or likes vs. does not like." (Hyman 1984, 67-68, cited in Hurford 1990)

This distinction can be rephrased as a difference between characteristics of the *users of language* and characteristics of *the purpose of language use*. Hurford (1990:96) makes a useful analogy with the design of a spade. Parts of the spade are clearly designed with the *purpose* of the spade in mind, the sharp metal blade for example. Other parts of the spade appear to be designed more for the *user*, such as its hand-sized handle and the length of its shaft. We can see that both aspects of the use of the spade have influenced its design — the spade's structure fits its function because of this.

It has been suggested (e.g. Hall 1992, 32) that the functional approach suffers from a lack of cohesion. This stems partly from the fact that the study of the purpose-based aspects of function and the user-based aspects of function belong to rather different research traditions in linguistics. In principle, however, I believe that this need not be the case. The distinction highlighted by Hyman and Hurford can be subsumed by a view that looks solely at the *process* of language use. All aspects of the spade's design can be explained by carefully examining the aspects of the digging process — the user of the spade and the purpose of the spade are unified in this act.

The various aspects of function utilised in the explanations of the last section might be similarly viewed as aspects of language processing. Givón (1985:189) argues that iconic tendencies in language result from the relative ease of processing forms which are "isomorphic to experience". The work of Sperber & Wilson (1986) in Relevance Theory also places a great deal of importance on processing effort in explaining pragmatic effects. The discourse features that DuBois (1987) appeals to must similarly have their ultimate explanation in terms of processing. For example, the reason that given entities are pronominalised is presumably related to the relative effort it takes for a hearer to recover the referent for a given vs. a new entity.

Although it looks as if there are a multitude of different ways in which language use can impact on universals, many of these can ultimately be reduced to pressures of processing language in real time. Processing here is a general term for *both* the act of parsing (i.e. mapping an acoustic wave onto a corresponding message and interpretation) and production (i.e. the mapping from communicative intention to articulation). A functional explanation for a language universal therefore is a statement of fit between that universal and the pressures of processing. For the functionalist, a universal is explained if it appears to be designed to ease processing. I do not claim to have shown that all functional explanations can be reduced to considerations of language processing, merely that this might be the case for many. However, the rest of this thesis will be restricted to explanations that appeal to pressures on production and perception of language. Another reason for this decision is that there are available *a priori* theories of language processing that have been compared with cross-linguistic evidence. This serves to deflect a common criticism of functional explanations (e.g. Lass 1980) — that they are constructed "after the event" in the sense that there tends to be an *ad hoc* search for functions that match the universals to be explained.

1.3 UG and universals

As mentioned earlier, the functional approach to explaining language universals contrasts sharply with the other major paradigm in modern linguistics. As Hall (1992:2) puts it, "much, perhaps most, recent work within the functional approach either explicitly or implicitly uses the Chomskyan paradigm as a point of departure or a point of contrast."[6] One of the purposes of this thesis, particularly chapter 5, is to show that this opposition is spurious at best, and rather damaging for the explanatory adequacy of both approaches.

This apparently opposing paradigm goes under a number of different names — *Chomskyan*, *generative*, *formal* and *innatist* (or *nativist*) — all of which are rather misleading. Firstly, just as with the functionalist approach, these terms suggest an unwarranted degree of coherence. There are currently several broad theoretical programs to which these labels could be applied. For example, *Principles and Parameters* (or *Government Binding theory*) (Chomsky 1981), the *Minimalist Program* (Marantz 1995),

[6]The syntactic (as opposed to phonological) bias of this thesis should be clear by this point. The following review ignores the corresponding tension between functional and generative approaches to phonology.

and *Optimality Theory* (e.g. Grimshaw 1993). All of these are Chomskyan in the sense of directly expanding on the basic suggestions of Chomsky's own work, but there is a great deal of diversity even here. None of the theories within these programs is strictly generative or formal (although formalisation is possible), but the name seems to have stuck from the early days of transformational grammar. There *are* formal theories of syntax around, however; *HPSG* (Pollard & Sag 1994) is the most popular at the moment. On the other hand, these theories could not really be called "Chomskyan".

1.3.1 Syntactic theory and universals

The final term in our list — innatist — is perhaps the most useful for our purposes. It refers to an underlying idea that, in achieving explanatory adequacy, a theory of syntax must be telling us something about the human brain. In particular, it tells us about properties of the brain that are biologically given as opposed to given by the environment. Syntactic theory, in the innatist sense, is a theory of the knowledge of language with which we are born. This is important because any innate component to our knowledge of language can be assumed to be shared by every member of our species.[7] If this is so, then we have a ready-made explanation for universal properties of languages (Hoekstra & Kooij 1988).

It seems then, that the innatist and functionalist approaches are inevitably in competition as explanations of language universals. It is important to realise, however, that the central question that each approach is attempting to answer is different. Simplifying the situation drastically, the difference can be characterised in terms of questions posed to and answers given by an imaginary functionalist, and an imaginary formalist:

The innatist approach

Central question "How are languages acquired from the degenerate data available to the child?"

[7]This is not *necessarily* the case, of course. It is possible that some degree of variation in innate knowledge of language may be uncovered.

. . . *answer* "A richly specified innate language acquisition device (LAD) in combination with the Primary Linguistic Data (PLD) is sufficient for the task."

Subsidiary question "Why are there constraints on cross-linguistic variation?"

. . . *answer* "The structure of the LAD constrains variation."

The functional-typological approach

Central question "Why are the constraints on variation of a particular form?"

. . . *answer* "The particular observed constraints are the reflex of language use."

Subsidiary question "How are languages acquired?"

. . . *answer* "The data available to the child is rich enough that language can be acquired using general purpose learning mechanisms."

The richly structured, innate "Universal Grammar" or "Language Acquisition Device" posited by generative syntax is not proposed in response to the hierarchical and parametric universals uncovered by typological research. Instead, the prime concern is the problem of language acquisition in the absence of necessary experience — a variant of *Plato's problem* in Chomsky's (1986) terms. A brief review of the solution given by the Principles and Parameters approach will make this clearer (for a more in-depth review, see e.g. Haegeman 1991, 10–20).[8]

1.3.2 Principles and parameters

Levels of adequacy

An interesting feature of the Chomskyan approach to linguistic theory is the recognition of two levels of adequacy of a theory. Firstly, a theory is *descriptively adequate* if it goes beyond a description of the linguistic data and instead accounts for a native speaker's intuitions. In order to do this it is essential to recognise that language has

[8] Recent developments in syntactic theory suggest a trend away from parametric theories of acquisition and variation. Instead, variation is being devolved to individual lexical entries. The idea of a core of invariant principles which constrain variation is still a central one, however.

two very different aspects: its external aspect and its internal aspect. External language (or E-language) is that aspect of language that is directly observable as writing or speech. Internal language (or I-language), on the other hand, is the specific knowledge of a person that allows her to produce or comprehend a particular language. I-language is therefore the domain of enquiry for a descriptively adequate theory of syntax, in the Chomskyan approach.

The preferred, though not sole, method of studying I-language is through careful elicitation of judgements of grammaticality. These judgements are assumed to abstract away from factors that influence E-language but have nothing to do with the internal knowledge of the speaker, such as processing constraints. This assumption underlies the *autonomy thesis*: the idea that I-language makes no reference to system-external factors (e.g. Chomsky 1975, cited in Newmeyer 1992, 783). This is perhaps another reason for the apparent opposition of formal and functional approaches. We will return to this issue in chapter 5.

The second level of adequacy of a theory of syntax — *explanatory adequacy* — is achieved if it can account for speakers' acquisition of the knowledge embodied in I-language. As noted above, the Chomskyan approach relies on the degeneracy of input data, the argument being that the acquisition of language can only be achieved given innate syntactic knowledge. Clearly, not all of language can be innately coded otherwise there would be no cross-linguistic variation. In Principles and Parameters theory, this variation is assumed to result from the setting of various parameters in response to the environment during acquisition. These parameter settings interact with an inventory of invariant principles which (in combination with a set of lexical items) make up the mature I-language of a speaker.

The contents of UG

Universal Grammar therefore has two properties (from Haegeman 1991, 14):

1. "UG contains a set of absolute universals, notions and principles which do not vary from one language to the next."

2. "There are language-specific properties which are not fully determined by UG but which vary cross-linguistically. For these properties a range of choices [parameters] is offered by UG."

The problem of language acquisition now boils down to the setting of parameters given appropriate triggering experience extracted from the PLD. Compared to the task of learning a language using some kind of general purpose learning mechanism, this parameter setting is relatively trivial. In this way, the Principles and Parameters approach appears to solve Plato's problem for language. Notice, however, that the very existence of this problem is not universally accepted:

> "How good is this argument? On the one hand, it seems to me highly plausible that there are *some* innately represented features of human language in the human species, and that these do facilitate language acquisition. On the other hand, there is a major issue that has not received the attention and critical scrutiny it deserves within the Chomskyan literature, namely: what exactly *can* the child infer from positive evidence? what kinds of learning strategies *do* children actually adopt, both in language and in other cognitive domains? and *are* these strategies systematically incapable of explaining language acquisition without the innateness hypothesis?" (Hawkins 1988:7)

Constraints on variation

Putting the learnability issue aside, what types of constraints on variation can this theory explain? The format of UG sketched above seems to directly allow for two types of universal. Firstly, the principles of grammar constrain languages in a formal or substantive sense. For example, the universal that languages allow sentential subjects is trivially predicted from the Extended Projection Principle which includes a requirement that clauses have a position for a subject.

Parametric universals also seem to be easily explained in this approach. The setting of a parameter to one "position" or another in the process of acquisition has typically many effects on the ultimate grammatical structure of the language. If this is the only (non-lexical) way in which languages can vary, and all other things are equal, then properties associated with a particular parameter setting should give rise to a parametric universal. So, for example, one parameter discussed by Haegeman (1991:450–451) determines the overtness of *wh*-movement in a language. English has overt *wh*-movement, whereas Chinese has non-overt *wh*-movement. The differences in the sentence structures of these two languages that this parameter difference creates could form the basis of a set of binary types which would then be related by a parametric universal.

Although it might seem counterintuitive given the nature of parameters, hierarchical universals can also be expressed in this theory. A multi-valued parameter (or a set of binary parameters) can, in principle, "point to" the position of a language on an implicational hierarchy. The possible *governing categories* in a language provide us with an example. These determine constraints on the positions of anaphors and their antecedents and appear to form a hierarchically ordered set. Manzini & Wexler (1987) propose a 5-valued parameter which inputs into a definition of a governing category:

> γ is a governing category for α if: γ is the minimal category that contains α and a governor for α and has either
>
> 1. a subject, or
> 2. an Infl, or
> 3. a tense, or
> 4. a "referential" tense, or
> 5. a "root" tense
>
> depending on the value of the paramater.

Now, the details of this definition and exactly how it affects the distribution of anaphors need not concern us here. The interesting feature of this definition is that different settings of the parameter give rise to different degrees to which anaphors may be separated from their antecedents. In fact, according to Manzini & Wexler

(1987), the grammatical domains within which anaphors and their antecedents can both occur form subset relations down the list of parameter settings above. In this way, hierarchical patterns of variation are expressible in Principles and Parameters theory.

A careful study of the typological correlates of parameters such as these is conspicuously absent from the literature and probably will remain that way. This is partly due to the gradual rejection of parametric variation in favour of lexical variation, and partly due to the nature of formal syntactic research, favouring as it does the in depth analysis of a few languages rather than the shallow analysis of many. Another reason why parameters do not readily translate as universals, however, is that their effects are highly interactive. The grammar of a language, and hence its resultant typological type(s), is a result of all the principles and parameter settings working together to constrain the set of grammatical sentences. If a particular observed universal is to be explained syntactically, it is not likely to involve one parameter but an examination of the possibilities allowed by the totality of UG.

Finally, whilst it is in principle possible that all the different logical forms of constraint described in this chapter can be expressed by a combination of parameters and principles, it is hard to see how this paradigm could be used to explain statistical universals. Of course, this is not its job (as pointed out in the previous section), but *at the very least* it leaves some scope for other forms of explanation.

1.4 The problem of linkage

The previous two sections have outlined quite different approaches to the problem of explaining language universals. I have suggested that both approaches eventually have their place in a complete view of universals. Although the full justification for this point of view must wait for later chapters, a basic flaw in each approach on its own should be pointed out here.

Firstly, although the innatist line of reasoning has many virtues — for example, it is explicit about the mechanism through which universals emerge — it fails to tackle the puzzle of fit. For example, the order of derivational and inflectional affixes could conceivably be constrained by some model of generative morphology. This constraint would then be assumed to be part of the biological endowment of the language learner, and would serve to partially alleviate the problem of learning language. As a side-effect, Greenberg's (1963) universal (28) would be explained. The problem with this is that it misses the fact that this universal *appears to be designed* with iconicity in mind. Our imaginary (extreme) nativist would have to assume that it was simply coincidence that the formal constraint happened to be iconic to "conceptual closeness" (Bybee 1985). So, perhaps this *is* a coincidence, or the theory of iconicity is sufficiently *ad hoc* in its formulation to be ignored. If, on the other hand, this fit of universal to processing can be demonstrated over and over again, this appears to undermine the innatist autonomy assumption (though, see chapter 5 for a different perspective).

The biggest flaw in the functional approach has already been mentioned. It highlights the fact that universals fit pressures imposed by language use, but this on its own does not constitute an explanation of anything. The innatist approach links universals to acquisition so that constraints on cross-linguistic variation are the *direct consequence* of constraints on the acquisition (and mental representation) of language. The functionalist approach fails to make this link between *explanans* and *explanandum* leaving the real puzzle, the puzzle of fit, unexplained. Bybee (1988:352) refers to this as the "how question" — given a set of generalisations about language she asks "how do such generalisations arise in language? What are the *mechanisms* that bring such a state of affairs about?" Hall (1988:323) argues that a proposed explanation must "attempt to establish the mechanism by which underlying pressure or pressures actually instantiate in language the structural pattern under investigation". The feeling that there is something missing from functional explanations is also echoed by Croft's (1993) complaint that linguistic theories of adaptation (i.e. fit) do not match up to biological ones:

> "... the sorts of explanations made by typologists are essentially adaptive ones: language structures are the way they are because of their adaptation to the function(s) of language ... In this respect linguistics also parallels

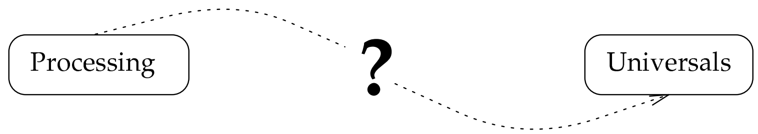


Figure 1.2. The problem of linkage. Compare this with the solution, figure 5.3.

> biology.
>
> However, the philosophical analogy between linguistic functional explanations and biological adaptation is not always fully worked out in linguistics." (Croft 1993)

To be completely explicit, we can formulate a *problem of linkage*:

> Given a set of observed constraints on cross-linguistic variation, and a corresponding pattern of functional preference, an explanation of this fit will solve the problem: how does the latter give rise to the former?

This thesis is an attempt to answer this question in a very general way (essentially to fill the gap in figure 1.2), but with examples from specific universals and specific theories of processing. As such, the main aim is not to uncover new constraints on variation, nor to find new functional asymmetries, although modelling the link between these two inevitably leads us to some new predictions both about universals and about processing.

In order to test that the proposed solution to the problem of linkage leads to the correct conclusions, I have adopted a simulation methodology. The theoretical assumptions of this thesis are therefore formalised as computer programs and tested against the available cross-linguistic evidence. This approach is fairly unusual in the linguistic literature, but it does have some precedents — for example, the evolutionary simulations of Hurford (1989) and other papers, Jules Levin's dialectology simulations reported by Keller (1994:100), and Bakker's (1994) computational work on typological theory testing in the Functional Grammar framework. The adoption of this methodology allows us to keep apart the general answer to the problem above from the specific examples of the explanatory approach (e.g. the accessibility hierarchy and Hawkins's (1994b) performance theory). The former is encoded as a simulation platform, and the latter as the particular initial conditions of a simulation run.

1.5 Overview

The rest of the thesis divides roughly into two parts. The first half goes into a theoretical approach to the problem of linkage and its computational modelling and testing on particular explanations in the literature. The latter half of the thesis then reflects on the implications of the proposed approach for typology, functional explanation and particularly innate theories of language variation.

The next chapter builds up a picture of the link between universals and function by considering in some detail Hawkins's (1994b) recent performance theory of word order universals. For this explanation to be complete, it is argued that the parser must be acting as a *selection mechanism* within the cycle of language acquisition and use. This view is shown to be related to characterisations of language change as an *invisible hand process* and to more general models of complex adaptive systems. Given this, a computational model of this system is built and tested using Hawkins' performance metric. It is shown that this model gives us a mechanism by which universals emerge, and as a bonus derives the prototypical S-shaped time course of change. The chapter ends with some discussion about the relationship of universals and *markedness* given this model.

Although the simulation seems to be successful at this stage, the types of universal on which it is tested are quite simple (e.g. two-valued parametric). Chapter 3 aims to extend the approach to explain the paradigm multi-valued implicational universal: the Accessibility Hierarchy. To do this certain changes need to be made to the model to allow for multiple stable types to coexist. Once again, Hawkins's (1994b) performance theory is applied to the task, but the initial results are disappointing. It is argued instead that Hawkins' explanation needs to be extended to a *competing motivations* approach in which speaker and hearer are in conflict in the acquisition/use cycle.

Two types of complexity are proposed which both input into the simulation; if these shift in relative prominence over time, the end result is a dynamic situation with the correct hierarchical pattern of linguistic variation moving geographically over time. This important result is explained using a simple graphical formalism based on graph-theory, and predictions are made and tested regarding more subtle distinctions in the strategies of relativisation available to speakers. Finally in this chapter suggestions are made for the extension of this approach to other hierarchical universals.

Having made the case for a selection-based solution to the problem of linkage, the focus changes in chapter 4 to the implications for the modes of explanation reviewed above. A failure in the functional approach is highlighted when other processing pressures on the comprehension of relative clauses are compared with the cross-linguistic evidence. Specifically a review of the psycholinguistic literature suggests that there is an asymmetrical processing preference for *parallel function* relatives. This appears not to be reflected in any language. There seems, therefore, to be something constraining the process of linguistic adaptation. It is argued that the best candidate for such a (meta-)constraint is an innate language faculty in the Chomskyan sense. This conclusion is strengthened by a careful examination of cases where parallel function apparently *is* expressed in languages. If the innate LAD can constrain the emergence of relative clause universals, it is probable that there will be other mismatches between form and function that can be similarly understood. The chapter ends with a look at animacy, length, Heavy NP shift and the English genitive in the light of this.

Chapter 5 takes the link between function and innateness one stage further with a review of the most recent literature on the biological evolution of the human language faculty. The very autonomous features of the LAD that appear to put its study in direct opposition to the functional enterprise are argued to have a type of functional explanation themselves. This means that the solution to the problem of linkage (the missing piece in figure 1.2) that was proposed in the first half of this thesis needs to be elaborated to take into account other forms of adaptation. A comparison of five different authors' views on the origin of the Subjacency Condition serves to highlight the lack of consensus in the literature on this subject.

Finally, in this necessarily speculative chapter and in the conclusion chapter 6, some suggestions are made about the directions future research might take, especially in the light of the approach taken in this thesis.

Chapter 2

The impact of selection on word order

In order to explore how pressures on language use can explain language universals, some theory of use must be put forward.[1] This chapter examines such a theory — the performance theory of John Hawkins (e.g.. Hawkins 1994a) — that has been mainly used to explain word order universals. Hawkins' theory provides us with an explicit quantification of the relative parsing complexity of various orders of terminal elements. The main thrust of this chapter will be to solve the problem of linkage in this specific case: how does a difference in parsing complexity lead to a difference in cross-linguistic distribution? Although this is a quite specific example of the fit of universals to processing, the solution will be developed in general terms and extended to other examples later in the thesis.

2.1 Hawkins' processing theory and word order

Hawkins' performance theory (Hawkins 1990, Hawkins 1992a, Hawkins 1992b, Hawkins 1992c, Hawkins 1994a) has been applied to two separate but related explanatory domains. On the one hand he examines the choice of word orders in performance both

[1]Some sections of this chapter have been published as Kirby 1994a.

within so-called fixed order constructions by rearrangement rules such as English heavy NP shift and in "free-order" constructions. The other area is the distribution of basic word-orders, grammaticalised in competence grammars across languages and it is this second domain — that of word order universals — that is the central concern of this chapter.

Two proposed explananda are:

Head ordering The *statistical* tendency for languages to have a consistent positioning of heads relative to non-heads across the phrasal categories in the competence grammar.

Left-right asymmetries Short constituents such as pronouns tend to appear to the left of heavy constituents such as relative clauses in competence grammars as well as in performance.

Hawkins uses a large sample of languages classified into types (Hawkins 1983) to demonstrate the validity of these empirical generalisations, expressing distributional universals as ratios of exemplifying languages to non-exemplifying languages (e.g.. there is a clear tendency for SOV languages to be postpositional – 93% in Hawkins sample). Matthew Dryer's work on word order universals (e.g.. Dryer 1991; Dryer 1992) goes further than Hawkins' since it takes into account the idea that simple language counts cannot be used to demonstrate *statistically significant* differences in numbers of languages, because statistical tests require items in a sample to be independent of each other. In order to meet the criteria of independence a language sample would need to consist of languages that were genetically and areally unrelated to each other. Consequently, any such sample would probably be too small to make any significant generalisations. I will return to Dryer's work later, but for now I would suggest simply that correlations as strong as SOV&Po, above, in a large sample are presumably significant without consideration of genetic/areal groupings.

2.1.1 The parser

Hawkins' main parsing principle, Early Immediate Constituent Recognition (or EIC), is expressed as a preference of the human parser for as much constituency information as possible in the shortest time. Hawkins argues for this preference with reference to the literature on parsing and also defines a method for quantifying this preference. This section summarises Hawkins' arguments which are treated more fully in Hawkins 1990.

Modules of mind In the dedication of "The Modularity of Mind" (Fodor 1983), Fodor quotes a comment made by Merrill Garrett that parsing is basically "a reflex". He argues that various modules of the mind dealing with input — including the parser[2] — have reflex-like properties. Some of these properties are:

Domain specificity Analysis of highly *eccentric* stimuli (such as acoustic waves organised into sentences) requires a set of information that is specific to the domain of those stimuli.

Mandatoriness The response of an input system to a stimulus provided by sensory transducers is obligatory — it is impossible not to attempt to parse a sentence, for example, if you hear it.

Encapsulation Input systems have only very limited access to high-level information in the form of expectations or beliefs. So, for example, it should be possible to parse a sentence without *necessarily* bringing higher-level knowledge into play in the parsing of that sentence.

Speed Input systems are surprisingly fast. This speed of operation is linked closely with mandatoriness: if an input system acts like a reflex, then computation can — indeed, must — begin immediately the stimulus is presented. Time is not

[2]Though I am treating the parser as one of Fodor's "input systems" it is possible that similar principles may play a part in the generation of output. The parser, therefore can be seen as *one* of the processing mechanisms mediating between the two parts of the Saussurean sign. It may turn out that processing considerations have a large part to play in the choice of orderings of sentences produced, but for the moment I will only be looking at the role they have in comprehension (see later).

wasted "making up our minds" about how to deal with the input, as Fodor puts it.

Hawkins uses these features of modules of mind to argue that the parser will construct hierarchical structure as rapidly as possible when given enough information to do so (mandatoriness and speed). These facts also suggest that a model of the parser should only rely on information specific to the parser, i.e.. a grammar, and feedback from other parts of the language system, such as pragmatic knowledge, should not be postulated (domain specificity and encapsulation). Frazier & Rayner (1988) give empirical support to this claim by comparing reading times of sentences with sentential subjects with those where the subject is extraposed (e.g.. *That both of the conjoined twins survived the operation is remarkable.* vs. *It is remarkable that both of the conjoined twins survived the operation.*) The difference in reading times between the pairs of sentences was similar whether they were presented in or out of a context that introduced the relevant referents. This suggests that non-syntactic information is not used to alleviate processing difficulty.

Deterministic parsing Another important feature of the human parser is determinism. The system modelling the human parser described by Marcus (1980) crucially relies on this feature:

The Determinism Hypothesis The syntax of any natural language can be parsed by a machine which operates "strictly deterministically" in that it does not simulate a non-deterministic machine. (Marcus 1980:§1.1)

The parser, then, will build a mother node above a syntactic category immediately and obligatorily as soon as its presence is guaranteed by the input and the phrase structure rules of the language. In general, this will occur whenever a syntactic category *uniquely determines* a mother node. These mother node constructing categories (MNCCs) are similar to heads in traditional syntactic theory, but may also include some closed-class function words such as determiners which uniquely construct noun phrases. So, for example, in the verb phrase, *tended the garden*, *tended* can construct VP, and *the* and *garden* can both construct NP. This gives us Hawkins' first parsing mechanism:

Mother Node Construction During parsing, if an MNCC is discovered, then the determined mother node is built above the constructing category immediately and obligatorily.

Other constituents that are immediately dominated by a mother node may be encountered before or after the MNCC. In either case they are attached to the mother node as rapidly as possible after it has been constructed:

IC Attachment Immediate constituents that are discovered before the MNCC for a particular mother node are placed in a look-ahead buffer for non-constructing nodes. As soon as a mother node is constructed, all ICs that can be attached to the mother node in accordance to phrase structure rules are attached as quickly as possible, either by removal from the buffer or by being encountered later in the parse.

The human parser must obviously use more than just these two parsing mechanisms, but these two will be enough to motivate the parsing principle, Early Immediate Constituent Recognition.

2.1.2 The EIC metric

Early Immediate Constituent Recognition (EIC) is the most important of Hawkins' parsing principles and provides a method of calculating a measure of parsing difficulty for a particular tree structure and a particular grammar. The basic idea behind the EIC is that of the Constituent Recognition Domain (CRD) of a particular node.

Constituent Recognition Domain The CRD for a node $\mathcal{N}$ is the ordered set of words in the string being parsed starting from the MNCC of the first IC of $\mathcal{N}$ on the left to the MNCC of the last IC of $\mathcal{N}$ on the right and including all intervening words.

It is possible to attach all daughter ICs to a mother node on the basis of a subset of the words dominated by that mother node. It is this subset that is described by the CRD. So, for example, in the sentence *Brian hid under the tree*, all the ICs of the verb phrase may be attached after the words *hid under* have been parsed, since *hid* will construct VP, and *under* will construct PP which is the last IC of the verb phrase. As we shall see in the next chapter, this concept of relevant subsets of structure can be generalised to other psycholinguistic operations. Given that the parser will prefer to completely recognise structure as rapidly as possible, it is logical to assume that there will be a preference for smaller subset structures — shorter CRDs. Notice that the definition of CRD makes no mention of the MNCC of the mother node itself. If this occurs at the right end of the string, then the daughter ICs, once constructed, will be placed in a look-ahead buffer as described above, and will be attached once the mother node is constructed at the end of the string — the concept of the CRD, therefore, holds wherever in the domain the mother node is actually constructed.

Evidence for the validity of CRD length as a measure of parsing complexity can be seen in Particle Movement in English. In sentences (2.1-2.3) below, the CRD of the verb phrase (marked by an underbrace) is lengthened as the length of the noun phrase increases. Sentence (2.4), however, has a short CRD since the noun phrase is the last daughter IC of the verb phrase and the determiner constructing the noun phrase marks the end of the CRD:

(2.1) $Florence\ _{VP}\left[\underbrace{looked\ _{NP}[the\ phone\ number]\ up}\right]$

(2.2) $Florence\ _{VP}\left[\underbrace{looked\ _{NP}[the\ phone\ number\ of\ her\ friend]\ up}\right]$

(2.3) $Florence\ _{VP}\left[\underbrace{looked\ _{NP}[the\ phone\ number\ of\ her\ friend\ Dougal\ who\ she\ wanted\ to\ speak\ to]\ up}\right]$

(2.4) $Florence\ _{VP}\left[\underbrace{looked\ up\ _{NP}[the}\ phone\ number\ of\ her\ friend\ Dougal\ who\ she\ wanted\ to\ speak\ to]\right]$

It is quite apparent that the acceptability of the sentences decreases as the length of the CRD increases. Hawkins (1994a) gives many more examples that suggest re-arrangement rules in various languages tend to work to decrease the length of the CRD.

A metric can be calculated to quantify this preference for short CRDs and also to differentiate between CRDs of the same length to give preference to the CRD that gives information about constituency earlier in the left-to-right parse of the sentence. This metric reflects the parser's preference for the "earliest possible temporal access to as much of the constituency information as possible" (Hawkins 1990:233).

> **The EIC Metric** — the average of the aggregate left-to-right IC-to-word ratios of all the CRDs in the sentence.

> **Aggregate Left-to-Right IC-to-Word Ratio** — the average of all IC-word ratios for each word in a particular CRD where the ratio for a word w_j in a CRD $[w_1\ w_2\ \dots\ w_n]$ dominated by an IC_i in a set of ICs $[\mathrm{IC}_1\ \mathrm{IC}_2\ \dots \mathrm{IC}_m]$ is $\frac{i}{j}$.

I will not go into details of how Hawkins arrived at this method of calculation; suffice to say it in some way captures numerically the preference of the parser for access to as much constituency information as possible as quickly as possible within a particular "parsing window" — the CRD. The purpose of this chapter is to examine what can be said about word order universals *given* this metric. A different research topic could be the testing of the validity of this metric *as a reflection of parsing preference*, but to keep within the scope of the chapter, I assume that Hawkins is correct on this point.

2.1.3 EIC and competence

The EIC metric can be used to make predictions about not only the re-arrangement rules that might occur in performance, but also the basic orders found in the competence grammar. If we assume that the pressure from the parser will influence the word orders of the world's languages, we might expect to find the EIC metric for a particular construction to be reflected in the number of languages that allow that construction. Hawkins calls this the **EIC Basic Order Prediction** (essentially, a statement of fit):

> "EIC predicts that, in the unmarked case, the basic orders assigned to the ICs of phrasal categories by grammatical rules or principles will be those that provide the most optimal left-to-right IC-to-word ratios; for basic orders whose ratios are not optimal (the marked case), then the lower the ratio, the fewer exemplifying languages there will be." (Hawkins 1990:236)

Perhaps the most important prediction that the EIC principle allows us to make is that languages which have consistent left or right branching in binary tree structures will be more frequent than those that have inconsistent orderings. In the sentences below, the aggregate (i.e. average) left-to-right ratio for the verb phrase is shown (each word's ratio is shown next to that word):

(2.5) $$Brian\ {}_{VP}\Big[\underbrace{hid_{\frac{1}{1}}\ {}_{PP}[under_{\frac{2}{2}}}_{aggregate\ ratio=1}\ the\ tree]\Big]$$

(2.6) $$Brian\ {}_{VP}\Big[{}_{PP}[the\ tree\ \underbrace{under_{\frac{1}{1}}]\ hid_{\frac{2}{2}}}_{aggregate\ ratio=1}\ \Big]$$

(2.7) $$Brian\ {}_{VP}\Big[{}_{PP}\underbrace{[under_{\frac{1}{1}}\ the_{\frac{1}{2}}\ tree_{\frac{1}{3}}]\ hid_{\frac{2}{4}}}_{aggregate\ ratio=0.58}\Big]$$

(2.8) $$Brian\ {}_{VP}\Big[\underbrace{hid_{\frac{1}{1}}\ {}_{PP}[the_{\frac{2}{2}}\ tree_{\frac{2}{3}}\ under_{\frac{2}{4}}]}_{aggregate\ ratio=0.79}\Big]$$

The verb phrases of sentences (2.5) and (2.6) both have optimal CRDs because the MNCCs of the two ICs occur together. In general, for any binary branching tree, the optimal ordering in terms of the EIC metric will be that which *consistently* places MNCCs to the right or left of the non-constructing constituent. Since the head of a phrase is always an MNCC for that phrase, this seems to provide an explanation for the tendency for consistent head ordering across languages. The left-to-right nature of the EIC metric also predicts an asymmetry in *sub-optimal* phrases. Sentence (2.8) has a higher metric than (2.7) reflecting the extremely low proportion of SOV languages that have prepositions.

This is just one example of how the EIC metric is reflected in the competence grammars of the world's languages. Many others have been investigated by Hawkins and his collaborators.

2.2 Selection and emergence

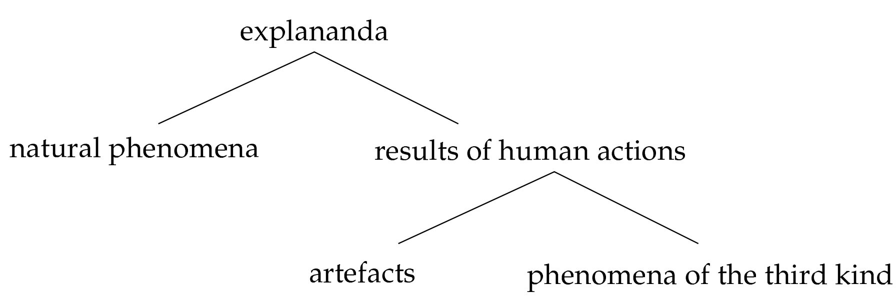
The explanation outlined in the previous section relies on an assumption — made explicit in the Basic Order Prediction — that parsing complexity is directly reflected in the distribution of types of grammars in the world's languages. A sceptical viewpoint on this assumption gives rise to the *problem of linkage* discussed in the last chapter. In this specific case, the problem of linkage is:

> How does a property of the human parser — namely the preference for early immediate constituent recognition — give rise to a restriction on the distribution of occurring languages in the space of possible languages — namely constraints on possible word orders in competence grammars?

To put it crudely, *even if* we have a theory of parsing that shows us that occurring languages are consistently less complex than non-occurring languages, we should still be puzzled and wonder, "how come the languages we find so neatly dovetail with the design of our parser?" The answer to this question relies on the idea that languages can *adapt*; this section argues that this adaptation is effected by a type of linguistic selection.

2.2.1 Universals are phenomena of the third kind

Keller (1994) puts forward an *invisible hand* account of language change. In this theory, language changes are viewed as *phenomena of the third kind*. Essentially, Keller gives us a typology of phenomena, dividing explananda into natural phenomena and results of human action, and further dividing the latter into artifacts and phenomena of the third kind.



These phenomena can be characterised as those "things which are the result of human actions but not the goal of their intentions" (Keller 1994:56). The process that gives rise to these phenomena is termed the 'invisible hand process'.

Keller discusses individual language changes as instances of objects of this kind. He gives as an example the change in the senses of the word *englisch* in German in the nineteenth century. In the early nineteenth century *englisch*$_1$ 'angelic' and *englisch*$_2$ 'English' were both used, but around the middle of the century the former disappeared. Keller points out that the explanation for this phenomena must refer to the actions of users of the language, and yet cannot be said to have been their *goal*. The explanation for the change involves setting out the *ecological conditions* that users of German found themselves in at the time of the change; *maxims of action* that describe the behaviour of individual language users; and the *invisible hand process* that gives rise to the non-local consequences of that behaviour (see Keller 1994, 93–95 for details of this explanation). The disappearance of *englisch*$_1$ in this view is an *emergent property* of the interaction of the users of German at the time.

Universals are similarly non-intentional results of human action. In other words, the local, individual actions of many speakers, hearers and acquirers of language across time and space conspire to produce non-local, universal patterns of variation. A description of the invisible hand process in this case is a theory of the propagation of variation through individuals. Indeed, the same mechanisms that explain individual language changes can be called upon to explain universals (although we are less interested in specific *ecological conditions*, as opposed to the universal pressures which will be relevant to each instance of change). A particular universal such as $SOV\&\neg Pr$ can be thought of as a *higher order emergent property*.

This brief discussion points to some desirable features we might look for in an explanation for universals. In particular, we should hope only to make reference to the actions of individuals at individual points in time. Furthermore, our model of the individual must describe precisely the relationship between these actions and the ecological conditions in which the individual is situated.

2.2.2 The Arena of Use

Figure 2.1 shows the cycle of language use and acquisition discussed in Hurford (1987:20–53). Here both the innate LAD and the "Arena of Use" are shown to play a

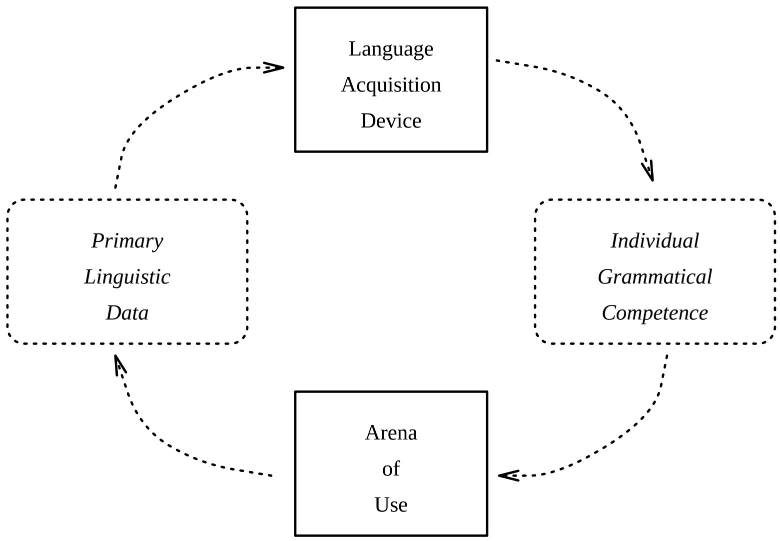


Figure 2.1. The augmented Chomskyan diagram for the linguistic cycle

part in *determining* language structure. Hurford (1990) describes the latter as follows:

> "The Arena of Use is where utterances . . . exist. The Arena of Use is a generalisation for theoretical purposes of all the possible non-grammatical aspects, physical, psychological, and social, of human linguistic interactions. Any particular set of temporal, spatial, performance-psychological and social coordinates for a human linguistic encounter is a point in the Arena of Use."(p.98)

> "As for the usefulness of coining the expression 'Arena of Use', my purpose is to focus attention on a vital link in the transmission of language from one generation to the next."(p.100)

Where should the parser, or other processing mechanisms be placed in this scheme? This depends crucially on a definition of "primary linguistic data". If PLD is taken to mean the linguistic data that the language learner hears, then the parser must sit on the arc between the PLD and the LAD. However to say that the PLD is *linguistic* data is begging the question: how does the child filter out other acoustic information, such as coughs, whistling or even foreign language sentences? Whatever the definition of PLD, some processing mechanism *must* exist in the arena of use to act as a filter. Some might argue that the LAD contains the necessary machinery to filter out non-linguistic data, but this explanation is unsatisfactory, since the same machinery must be used even after acquisition ceases, suggesting that it must be a separate module. The strong definition of primary linguistic data that I put forward is therefore *the data that a child attends to as linguistically salient*. All innate processing mechanisms can be distinguished from the LAD by the fact that they deal with a superset of the primary linguistic data. This superset of "raw" data is filtered by the processing mechanisms to provide the primary linguistic data for the language acquisition device. In fact, in order to dispel confusion, we might dispense with the term 'PLD' altogether and simply refer to *language data* and *trigger experience* for pre- and post-filtering linguistic data respectively. Lightfoot (1989) makes precisely this point in connection with learnability theory:

> "The trigger is something less than the total linguistic experience . . . the child might even be exposed to significant quantities of linguistic material that does not act as a trigger . . . "(p.324)

> "This means that children sometimes hear a form which does not trigger some grammatical device for incorporating this form in their grammar. Thus, even though they have been exposed to the form, it does not occur in mature speech."(p.325)

Interestingly, arguing from a connectionist viewpoint, Elman (1991) also suggests that the trigger experience will be a subset of the total raw linguistic data. He shows that, for a connectionist model to successfully learn a non-trivial grammar, the data used for "acquisition" must be presented in stages from simple to more complex. Consequently, his model initially incorporates a memory limitation which effectively filters out the more complex grammatical structures.

There are other logically possible means by which a parsing preference might make itself felt in the acquisition/use cycle. One could hypothesise that the human language *generation* mechanisms are subject to similar considerations of syntactic weight as the parser and thus that the generation of sentences that are difficult to produce will be avoided. The nature of human language generation is relatively poorly understood, but it has been suggested (e.g. Hawkins 1992c) that speakers may respond to considerations of parsing efficiency since the primary goal of the speaker is to effectively communicate to a hearer. Hence the production of sentences that are hard to parse is avoided specifically for the reason that they will be difficult to understand. However, if the parser filters sentences from the acquisition/use cycle, then it is unnecessary to postulate this kind of speaker/hearer symmetry in order to model the influence of the parser on language change. These issues will be discussed later in this chapter.

2.2.3 Complex adaptive systems

Gell-Mann (1992) suggests that language change can be characterised as a complex adaptive system — a system that can evolve. A complex adaptive system is a system which can compress information from the environment into a set of rules or schemata. These rules can then 'unfold' to produce effects in the environment which in turn may become input to the system (see figure 2.2 from Gell-Mann 1992, 11). In biological evolution, for example, the DNA of an individual is a compressed schema; this schema unfolds during the development of an individual and produces effects in the environment.

Now, in general these systems are *adaptive* because there is competition amongst schemata. Whether a particular schema survives this competition will depend on the viability of that schema in the environment. We say that the schemata are *selected*, and the models of this process are *selection models*. In biological terms, the survival of a sequence of DNA — or rather the information encapsulated in the sequence, since the particular physical embodiment of the information is temporary — depends on the

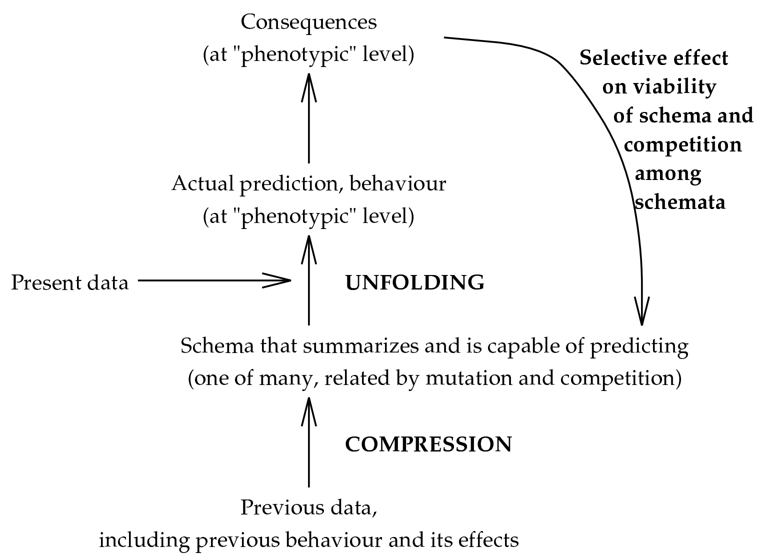


Figure 2.2. The operation of a complex adaptive system.

ability of the individual with that DNA sequence to survive and reproduce. Furthermore, in an environment with finite resources, the individual will be in competition with other individuals with other DNA sequences. In this way, adaptive systems tend to display an "apparent design" or "dovetailing" with the environment. This is just what we see with language universals which seem to show languages' structural properties dovetailing with the needs of the users of the language. As mentioned in chapter 1, good examples of this can be found by looking at implicational hierarchies. For example, the Keenan & Comrie (1977) accessibility hierarchy essentially states that, for a given number of types of relativisable noun-phrases, the specific types a particular language will allow are those that present the least difficulty for the users of the language to process (given certain models of parsing). This example is examined in depth in the next chapter.

Biological evolution, human cultural evolution, global economics and individual learning have all been studied as complex adaptive systems. How might this paradigm be applied to *glossogenetic* language evolution (i.e. language evolution on the historical timescale)? The essential features of the system — compression of regularities

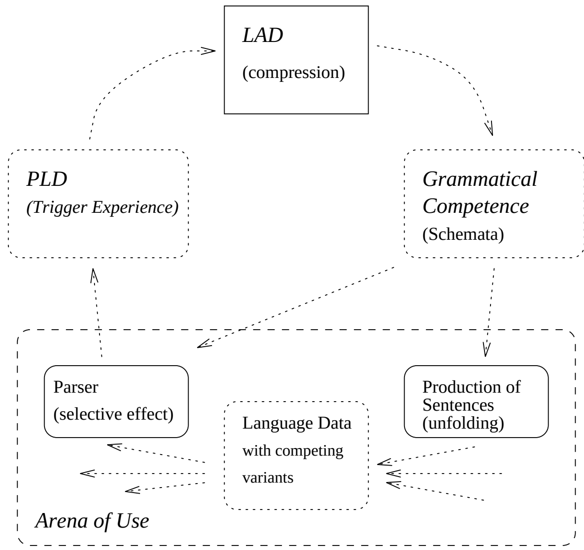


Figure 2.3. The acquisition/use cycle as a complex adaptive system.

in the data into schemata, unfolding of schemata to produce new data, and selective pressure on competing schemata — are all shown to be features of the language acquisition/use cycle in figure 2.3. Grammatical competence contains rules/parameter settings/lexical entries (schemata) of some kind that express, in a highly compressed form, regularities in the trigger experience. These unfold to produce utterances in response to features in the environment. Some of these utterances may then be filtered by the parser from the linguistic data input to the trigger for the next generation, providing a selective effect on the viability of the schemata that produced them. This selective effect is related to parsing principles such as EIC.

If the parser is a filter between raw data and the trigger experience, then it is possible that only some of the orderings of a particular constituent that occur in the raw data will be acquired. In order for Hawkins' explanation to work in this context, the probability of a particular utterance being used for acquisition will be proportional in some way to its EIC metric. It is possible, then, that different orderings in performance can become *fixed* in the competence grammar, or in a less extreme case, different orderings may become marked in some way.[3] The generalisation is that *in the process of acquisition* the EIC metric may make itself felt by influencing the variability of word orders that the child learns. This argument is equivalent to one that claims that acquisition and language change are dependent on text frequency. This from a recent paper on computational modelling of parameter acquisition:

> "If . . . a parameter is not expressed frequently in the input text, the learner will be under less pressure to set that parameter in accordance with the target setting. In this case . . . either the correct setting or the incorrect setting can survive in the linguistic environment." (Clark & Roberts 1993:301)

The only modification here is to view the "input text" as the input to acquisition *after* parsing.

It is likely that a particular ordering will not disappear suddenly from a language, so a sensible assumption is that the EIC metric changes the *frequency of use* of a particular ordering through the process described above. This seems to suggest that the child must learn, not only a particular construction, but a frequency as well. However, this assumption is not necessary for a description of gradual language change, if we define frequency of use of an ordering as being a reflection of a particular speech community's use of that ordering. In other words it is possible to have different frequencies for different orders without compromising a theory of "all-or-nothing" competence. The frequency of use of a particular ordering by one generation is some function of the frequency of use of that ordering by the previous generation and the EIC metric of that ordering. I shall refer to this process whereby a particular word order pattern gradually becomes fixed in the competence grammar as *grammaticalisation*. This term has been used by a large number of scholars to refer to diverse linguistic phenomena (see for example Heine *et al.* 1991). Traugott & Heine (1991:1), however, admit the use of the term in this case by defining it as "that part of the theory of language that focuses

[3]There is a general problem of circularity involved in any filtering of the PLD that appeals to grammatical competence. Since the parser must make use of a competence grammar *in order* to provide input to the acquisition process, it is pertinent to ask how such a competence ever arises. Jim Hurford (personal communication) has suggested that this circularity can be avoided if acquisition is looked at *incrementally* in stages from primitive structures to more sophisticated.

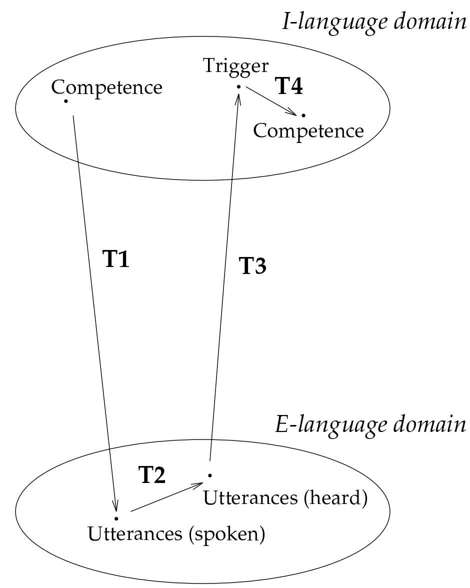


Figure 2.4. Transformations within and between *I*- and E-domains.

on the interdependence of langue and parole, of the categorial and less categorial, of the fixed and less fixed in language."

2.2.4 Linguistic selection as transformation

To recap on the ground we have covered so far: the desirable features of an explanation that appeals to use have been set out by characterising the explanation in terms of the invisible hand, and it has been argued that the influence of processing on language competence should be seen as a selective influence. More properly, functional pressures must influence the selection of linguistic variants that are *competing* in some way, and this selection must occur at some point in the cycle of language acquisition and use. Another way of seeing this is that there is a transformation that maps the competence of a speaker at some point in time to the competence of a speaker in the same speech community at some later time. Functional selection influences this transformation in a predictable, though statistical, manner.

Viewing linguistic evolution in terms of laws of transformation closely parallels biological thinking (as we should expect given that selection theories are general). So closely, in fact, that we can usefully borrow a map of transformations from Lewontin (1974) (cited in Sober 1984), replacing *genotypes* with *I-language* and *phenotypes* with *E-language*. The first important feature to note about figure 2.4 is that the transformation from competence to competence involves objects in two very different domains. The *I-language* domain contains objects in individual speakers' brains. The objects, the domain in which they exist and the transformation **T4** (acquisition), are what Chomsky (1986) argues are the proper target of study in linguistics.

On the other hand we have the *E-language* domain which contains *utterances* in some broad sense. These objects are more ephemeral, and are typically viewed as epiphenomena in the Chomskyan program. The transformation **T2** involves features of the world at particular points in time, for example, the level of noise, the availability of hearers, and so on.

Finally, we have the transformations **T1** and **T3** which map objects in one domain to those in the other. The former is mediated by speakers (production), and the latter by hearers (parsing). Both these transformations and those that map between objects within domains are not well understood by linguistic theory, but it is generally assumed that some innate (and therefore universally shared) neurological mechanisms play a role. In particular the focus of the explanation in this chapter is on the role of complexity of processing in influencing the transformation **T3** — in other words, the effect of EIC.

2.2.5 Replacement through competition and the notion of fitness

Given a simple case of two linguistic forms in competition somehow and the model outlined above, what might we expect to happen? Kroch (1989a); Kroch (1989b) discusses the rise of periphrastic 'do' in English as a case of replacement of one form with another, so a brief review of his work will be useful in this context.

Firstly, some terminology: given a linguistic form f carrying out some function $\mathcal{M}$, f' is a *variant* form of f carrying out the same function $\mathcal{M}$. The variants f and f' will typically occur as doublets historically and will be viewed of as synonymous to native speakers.[4] Finally I will use $f > f'$ to mean that f is preferred in performance for some reason, and F or F' to signify a language type in which the form f or f' is *basic*.

Kroch (1994) refers to situations where languages change their relative frequency of variants as *grammar competition*. Under his formulation, two or more broadly syntactic doublets behave in the same way as morphological doublets in competition for a paradigm slot. This view follows from a move in syntactic analysis to treat cross-linguistic variation as a reflection of variation in the properties and inventories of functional heads.[5]

> "If we take this view seriously, we are led to the conclusion that syntactic variation should be governed by the same principles as variation in morphology, since the locus of the variability in the two cases is the same — the formative. Just as morphological variants which are not functionally distinguished are disallowed, so we should not expect to find variation between semantically non-distinct syntactic heads. To the extent that such variability is found, it poses the same theoretical problem as the appearance of doublets does in morphology." (Kroch 1994:5)

Kroch points out that the "blocking effect" in morphology (whereby the presence of an irregular form in a paradigm slot blocks the occupation of that slot by a regular form) is a central tenet of modern morphology. However, doublets *are* in fact often observed in languages. However, if the doublets are functionally equivalent, speakers "learn either one or the other form in the course of basic language acquisition, but not both" (p. 6). Later on the same speakers may recognise the existence of the variant form, which "for them has the status of a foreign accent" (p. 6). Finally, one of these two

[4]It is likely that these sorts of truly synonymous variant forms are actually uncommon if they occur at all. Instead, variants will belong to a gradient scale of functional differentiation. This is a complex issue to which we will return in chapter 4.

[5]Notice that there is considerable possibility for confusion here. There have now been three different senses of 'functional' used. Firstly, 'functional' in the sense of 'to do with the functions of language in discourse and communication' as in *functional explanation* or *functional load*; secondly, 'functional' in the sense of 'carrying out some internal *linguistic* function' as in *functional differentiation*; and finally, 'functional' in the sense of 'belonging to the set of grammatical, closed-class morphs' as in *functional head*, here.

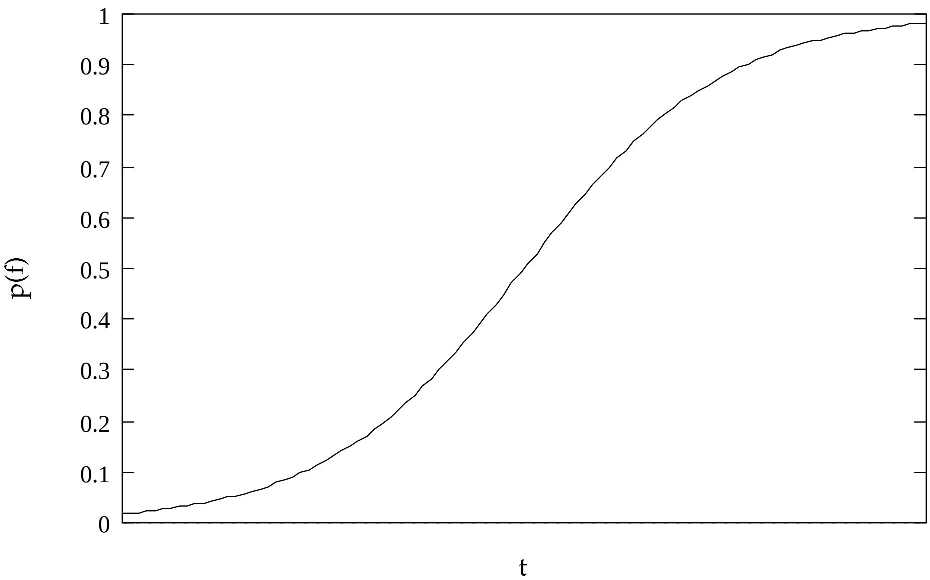


Figure 2.5. The time course of replacement through competition

doublets will tend to win out in a particular community — thus justifying our use of the term *competing variants* — or the two forms will become functionally differentiated.

Now, given the doublet forms f and f' where $f > f'$, we would expect the frequency of f in a speech community to increase over time. What would the time course of such a change look like? Well, a simple mathematical model of the replacement of forms through competition is available (Kroch 1989b):

$$p(f) = \frac{e^{k+st}}{1 + e^{k+st}}$$

where t is time, k is a constant determined by the initial frequency of f and s is the slope parameter, related to the degree to which f is preferred to f' (see figure 2.5). The shape of this function makes sense intuitively if one realises that the rate of growth of a new form is related not only to the numbers of that form already about, but also to the number of forms to be replaced (the derivative of the function above is $sp(f)p(f'))$. So, the slope of left hand of the graph in 2.5 is shallow since there are few fs about, and the right hand is shallow since there are few f's left to replace. It is suggestive in the light of arguments in the previous section that the same *logistic* function is used in biology to map the replacement in a population of genetic alleles that differ in Darwinian fitness (Spiess 1989, cited in Kroch 1989b). The fit of observed *syntactic* changes to this function has been tested by Kroch (1989b) and shown to be good. This lends further weight to the suggestion above, that syntactic as well as morphological change proceeds through a process of replacement by competition.

To recap, by treating word order variation as something arising from properties of individual functional heads, we can argue on theoretical grounds that a blocking effect similar to that in morphology is expected. Given differences in functional preferences, this leads us to expect change to follow from replacement by competition and that the time course of the change can be predicted by the logistic function.

The next question that must be addressed is how to fit a performance theory like Hawkins' into a model of replacement through competition. Given the abstract example above and all else being equal, a language of type F' will change over time into a language of type F. The manner in which f forms win out is by selection in **T3** because $f > f'$ — in other words, f is preferred to f' in parsing. In general, we can define this preference in terms of *fitness*, where *fitness* is a function from frequency ratios of pairs of variants to the average probability of acquisition of those variants. Where there are only the two variants under consideration, a plot of fitness by variant frequency gives us a simple graph with fitness increasing as the frequency of f increases. For reasons that will become clear in the next section, this graph is referred to as a *fitness landscape* and languages, according to this theory, will tend to 'climb' these landscapes and maximise fitness, in other words, through a process of selection, they will organise themselves to maximise the chances that their variants will survive in the Arena of Use.[6]

It might now be clear that the role of a theory of parsing complexity such as John Hawkins' is to provide a description of fitness landscapes. This conception of functional pressures — the first step to solving the problem of linkage — will be

[6]Chapter 5 discusses exceptions to the general rule that languages will maximise their own fitness.

useful in understanding the behaviour of a computer simulation of linguistic selection described in the next section.

2.3 A simulation of the complex adaptive system

In order to understand the implications of the model introduced in the previous sections and to ensure that its details are completely explicit, computational simulations of the adaptive process can be constructed. These simulations give us a way of experimenting with theory in some sense. A simulation of a theory in combination with a certain set of initial conditions can be used to see if the implications of the theory that we expect actually hold. Each run of the simulation can be seen as an experiment — not with real languages or real language users, but with virtual languages and virtual users whose relevant characteristics are defined by the way in which the simulation is set up. In the case of complex adaptive systems, the use of computer simulation is particularly appealing since emergent properties are expected to occur when many interacting virtual users are brought together: properties whose appearance may be hard to predict analytically. This is especially true of the simulations introduced in the next chapter. However, this section introduces a simple simulation of the linguistic selection of competing variants, and shows how the initial conditions can be set up which give rise to a behaviour characterised by the curve in figure 2.5.

2.3.1 Components of the simulation

The simulation system which underlies the results in this chapter has the following simple components which directly correspond to parts of the model described above:

Utterances These are the E-domain objects in figure 2.4. In the simulations described in this chapter these utterances are not actual sentences, but simply types or features of sentences. So, for example, an utterance in this sense could be *SVO*, or +*coronal* depending on what was being investigated.

Arena of Use This is an unstructured pool of utterances in these simulations (though, see chapter 3).

Grammars These are the I-domain objects in figure 2.4. In the simulations in this chapter they are simply lists of possible *utterances*. A typical simple grammar might be *[SOV, NAdj]*. This is one possible idealisation; another possible approach would involve the use of *parameters* to model I-language (see Niyogi & Berwick 1995).[7]

Speakers The simulations start with a speech community which is made up of a set of speakers each of which consists of a grammar. These grammars produce utterances for input to the Arena in the way described below.

Acquirers These are speakers who have yet to be assigned grammars. They take input as utterances from the Arena as described below.

It should already be obvious that the basic components of the simulation are gross idealisations of their real-world counterparts. This is just as it should be, however. Remember that the purpose of the simulation is not to be a complete analogue of the real world, rather it should be a reification of a theory. It should involve all and only the idealisations that a model of that theory would involve. If we were to build a simulation of some theory of the flocking of birds, let us say, and we built in a detailed description of wind direction which the theory did not mention, then the results of the simulation tell us nothing about the validity of our original theory. Of course, the process of building and testing the simulation might lead us to conclude that the original theory does not work *without* taking into account wind direction but

[7]Niyogi & Berwick's (1995) recent paper analyses the dynamics of a system involving parametric variation. In particular they derive the S-shaped curve (a result identical to that independently arrived at by the simulation in this chapter). The main difference between their model and ours is that they do not assume that a probability distribution is imposed externally by linguistic selection. In the simplest case (involving one binary parameter), the change in two possible grammars is determined by the different distributions of sentences in the grammars' output which trigger either setting of the parameter. In this way Niyogi and Berwick appeal to features *internal to the I-domain* to derive the time-course of change (see also the discussion in chapter 4, for further examples of the importance of the I-domain). A fascinating and important research project would combine Niyogi and Berwick's approach to parametric change, and the approach of this thesis to the fit of universals and processing. Sadly, this is rather beyond the scope of this thesis.

this simply serves to underscore the importance of simulation. Throughout this thesis there will be several cases where a theory will be shown to be inadequate through simulation in such a way.

The components listed above interact in the simulation according to two dynamic processes:

Production Speakers add utterances to the Arena of Use in line with their grammars. For the simulations in this chapter this is done completely randomly.

Parsing/Acquisition Acquirers become endowed with a grammar (and thus become speakers) in the following way:

1. A random subset of utterances in the Arena of Use is taken and forms each acquirer's linguistic data.
2. This subset is then modified through a process of filtering to form a trigger experience. The process of filtering involves measuring the distribution of utterances in the linguistic data, and then choosing from those utterances in such a way that the probability of an utterance appearing in the trigger is related to its distribution *and to its pre-defined parsing complexity*.
3. The trigger is then directly mapped onto a grammar (i.e. if an utterance appears in the trigger, then it is added to the grammar).

A run of the simulation involves each speaker in the community producing some number of utterances to add to the Arena, and then each acquirer parsing/acquiring utterances from the Arena. The number of acquirers and speakers is always the same, so that after acquisition, all the old speakers are discarded (as is the Arena of Use) and the acquirers become the new speakers for another iteration of the process.

2.3.2 A simple simulation: two competing variants

The details of the set up of the simulation depend on the particular feature of interest. A simple simulation should make the process clearer. Firstly, imagine a language with basic VO order and postpositions. According to Hawkins such a language would suffer from a sub-optimal EIC metric in structures such as $_{VP}[V\ _{PP}[NP\ P]]$ since the CRD for the VP stretches across the NP. Now, if a minor variant — prepositions — were introduced into that language, perhaps through language contact, then we would expect it to be preferentially selected from the Arena of Use by hearers because of its improved EIC metric. As a result, we would expect the frequency of prepositions in the Arena to increase over time.

To test this with the simulation, the initial speech community is made up of 450 speakers with the grammar *[postp]* and 50 speakers with the grammar *[prep]*. The manner in which each acquirer's trigger is selected from the linguistic data sampled from the arena[8] is governed by the following equations:

$$p(prep) = \frac{1.n_{prep}}{1.n_{prep} + 0.79.n_{postp}}$$

$$p(postp) = \frac{0.79.n_{postp}}{0.79.n_{postp} + 1.n_{prep}}$$

where $p(f)$ is the probability of the form f occurring in the trigger, n_f is the number of f forms in the linguistic data, and the values 1 and 0.79 correspond to EIC values for VO languages assuming a two word NP (Hawkins 1990:238). Any increase in the length of the NP would reduce the value of the postpositional EIC value making any effect of the dispreference more marked. Notice that these equations are the *fitness functions* for adpositions in VO languages.

The simulation was then run for 35 iterations after which the arena of use consisted almost entirely of prepositions, the originally minor variant. The graph in figure 2.6 shows the time course of the change (the vertical axis indicates the probability of finding a speaker with the grammar *[prep]*). The striking feature of this graph is its similarity to the S-curve (figure 2.5) which Kroch found in the historical data.

[8] The size of the sample was set to 30 utterances, which is large enough to give a fair approximation of the distribution of utterances in the arena without overly slowing the simulation down. The number of utterances allowed to each speaker is immaterial, since each speaker has only one possible utterance. This follows since acquirers in this simulation could only acquire one variant.

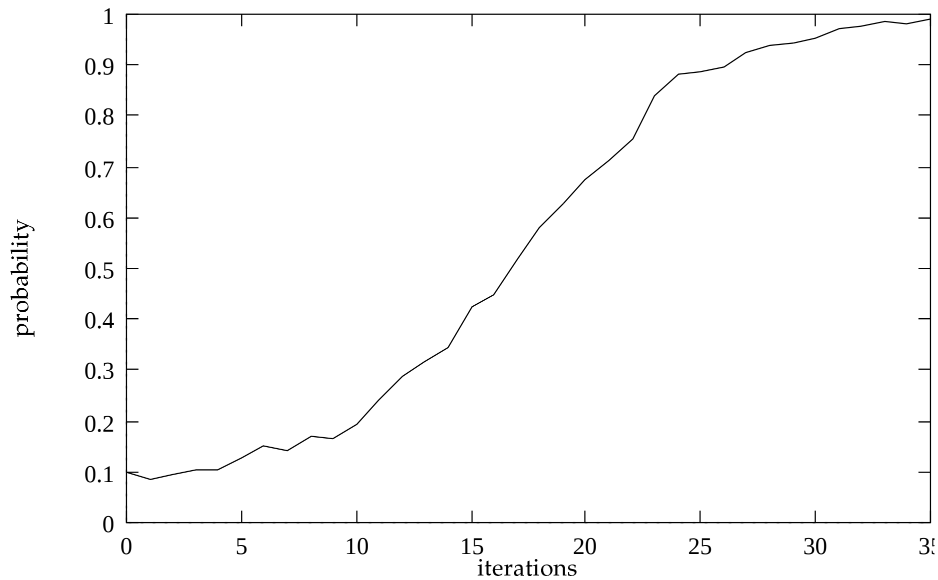


Figure 2.6. A simple run of the simulation showing VO&Postp changing to VO&Prep.

2.3.3 A previous simulation system

The simulation described here has its roots in the one discussed in Kirby 1994a which was applied to essentially the same problems. There are fundamental differences, however, which are worth highlighting.

Kirby (1994a:197) covers the basic elements of the previous simulation. The first obvious difference is in the grammar. In the earlier work this is a phrase structure grammar fragment rather than a list of forms. The grammar fragment is separated into two parts: immediate dominance rules and linear precedence *probabilities*. There is only one grammar in the simulation which is meant to model the behaviour of the whole speech community. In other words there is no explicit modelling of speakers or hearer/acquirers in the simulation.

The dynamic processes in Kirby's (1994a) simulation are also radically different from those used in the current work. Production involves randomly selecting an utterance allowed by the probabilistic grammar (with some arbitrary limit on recursive depth). This means that a much larger range of utterances is *in principle* possible than with the simple grammars used here. In practice, however, the grammars used for testing Hawkins' theory in the old simulation are made to be very simple so that it is clear what the simulation is doing, and the range of utterances kept to a minimum. Parsing/acquisition means taking each produced utterance and calculating its EIC value. If this value passes some *acquisition test* then the grammar is adjusted so that each ordering evinced by the utterance has its probability increased, at the expense of the other orders. The "acquisition test" is more likely to succeed if the EIC value is high *relative to the maximum and minimum of recent* EIC *values*.

The problem with this approach is that its relationship with the selection model is rather indirect. Individual speakers and hearers are not modelled, and the way in which the "acquisition" of an utterance affects the speech community (i.e. the linear precedence probabilities) is described by an unmotivated function. This means that nothing can be said about the time course of the changes, only about general trends for increase of one form over another. The reason given for these idealisations in the previous paper is that direct modelling of speakers and hearers would tend to lead to heterogeneous speech communities. In other words, the end result is expected to be a variety of language types. In fact, as we have seen, if the speech community shares one Arena of Use, variation appears to decrease. The following chapter explores these issues further by building structure into the Arena.

The other obvious difference between the two approaches is that the simulation in Kirby 1994a calculates EIC metrics on line, whereas the simulations here rely on a precalculation of the EIC preference. I submit that nothing hangs on this except for the time taken to run an experiment.

2.4 The model in action

This section describes three further trials of the simulation which lend support to the theory put forward in this chapter. All the examples are adapted from Hawkins 1990 and Hawkins 1994a and thus show how Hawkins' parsing theory in combination with a selection model of linguistic dynamics can explain the adaptive nature of various word-order universals.

2.4.1 Climbing a fitness landscape

In the example run of the simulation where a speech community adopted VO&Prep over VO&Postp, a function was described that mapped from relative frequency of adposition type to fitness. As the proportion of prepositions increases, then so the average fitness of forms in the speech community increases. This fitness is simply an average of the probabilities of each form surviving to the Arena of Use at the next iteration. The process of adaptation through linguistic selection acts to maximise this fitness.

Now, consider a situation where the language of the speech community could vary along two dimensions, rather than one: for example, adposition order and verb-object order. This involves a modification to the grammars of the speakers in the simulation which may be either *[VO, Prep]*, *[VO, Postp]*, *[OV, Prep]* or *[OV, Postp]*. The state of the speech community at any one time can be expressed as a point in a 2-dimensional space whose axes are the relative proportions of verb-object variants and adpositional variants. The interesting feature of this example is the way in which the fitnesses of the variants are related to each other.

The optimal orders in terms of parsing will be ones in which the heads (or, more correctly, MNCCs) are on the same side of their respective complements: in other words VO&Prep and OV&Postp. This is indeed what we find to be the most common orders in the world's languages. The parsing preference for prepositions over postpositions, then, is not absolute, but relative to the proportion of VO over OV in the Arena of Use, and vice versa. This co-dependent relation is modelled by the functions that filter forms from the linguistic data for the trigger experience:

$$p(prep) = \frac{w_{prep} n_{prep}}{w_{prep} n_{prep} + w_{postp} n_{postp}}$$

$$p(postp) = \frac{w_{postp} n_{postp}}{w_{postp} n_{postp} + w_{prep} n_{prep}}$$

$$p(vo) = \frac{w_{vo}n_{vo}}{w_{vo}n_{vo} + w_{ov} + n_{ov}}$$

$$p(ov) = \frac{w_{ov}n_{ov}}{w_{ov}n_{ov} + w_{vo} + n_{vo}}$$

where

$$w_{prep} = \alpha n_{vo} + (1-\alpha)n_{ov}$$

$$w_{postp} = \alpha n_{ov} + (1-\alpha)n_{vo}$$

$$w_{vo} = \alpha n_{prep} + (1-\alpha)n_{postp}$$

$$w_{ov} = \alpha n_{postp} + (1-\alpha)n_{prep}$$

and α is some constant showing the relatedness of the two variant pairs, with $\alpha > 0.5$ signifying that prepositions and VO are positively correlated, and $\alpha < 0.5$ signifying that postpositions and VO are positively correlated.[9] The actual value of α will depend on the average length of noun phrases in the utterances spoken. For the simulation runs in this section $\alpha = 0.6$.

The simulation was run eight times; each run started with a population of 500 speakers, with mostly (i.e. 90%) grammars that are uncommon in the world's languages. For half of the runs, the speakers mainly had the grammar *[VO, Postp]* and for half the runs *[OV, Prep]*. A plot of these runs is shown in figure 2.7. The results are non-deterministic in that the language of the speech community ends up either being one of the common cross-linguistic types, VO&Prep or OV&Postp, whatever the initial conditions.

We can see what is going on in this example by overlaying one of these runs on a plot of the function (for $\alpha = 0.6$):

$$F = \frac{w_{prep}n_{prep} + w_{postp}n_{postp} + w_{vo}n_{vo} + w_{ov}n_{ov}}{n_{prep} + n_{postp} + n_{vo} + n_{ov}}$$

[9] Notice that this assumes that the situation is symmetrical. In other words that the preferred types VO&Prep and OV&Postp are equally preferred, and that the dispreferred types OV&Prep and VO&Postp are equally dispreferred. However, the EIC metric is not symmetrical in this case: VO&Postp is preferred to OV&Prep (Hawkins 1990:238–239). The implications of this are explored in the next chapter.

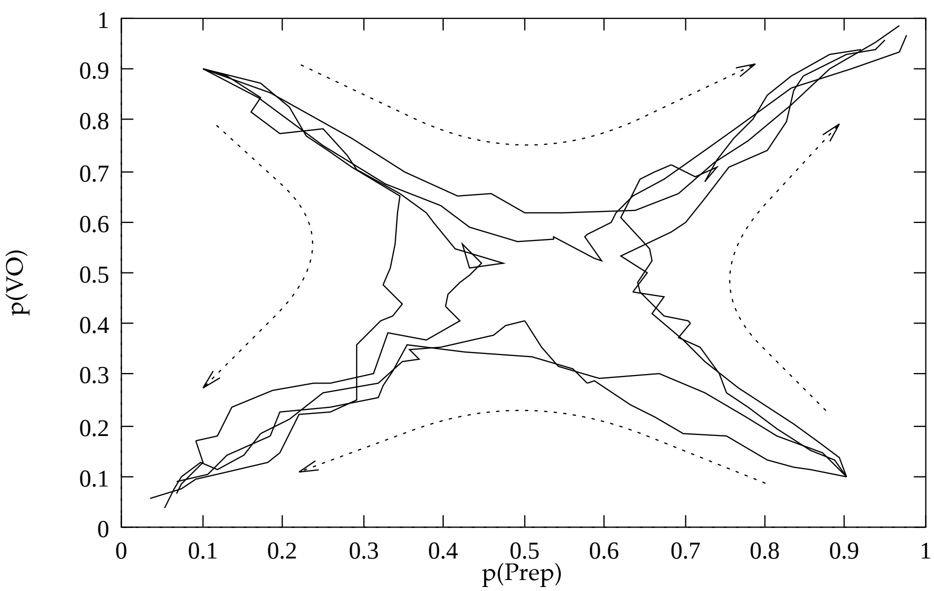


Figure 2.7. Eight runs of the simulation overlaid.

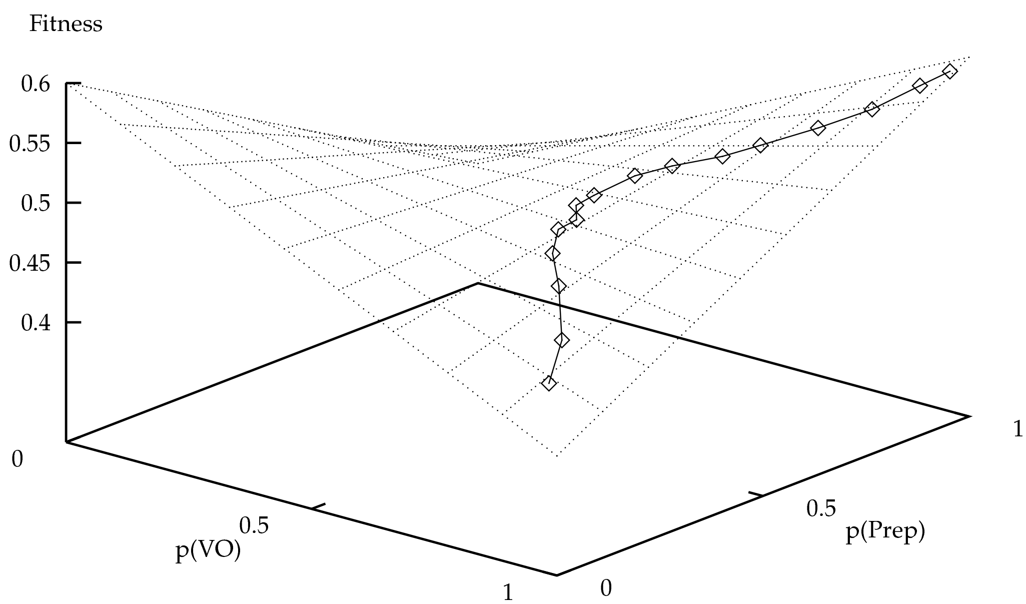


Figure 2.8. The simulation climbing a fitness landscape.

This is the fitness function for the example (i.e. the average probability of acquisition of variants in the speech community). The result is shown in figure 2.8. It is clear from this figure that the simulation is climbing the fitness landscape.[10] The important point of this graph is that the peaks of the fitness landscape correspond to common cross-linguistic language types; the fitness landscape is described by a theory of parsing complexity; and speech communities climb fitness landscapes through a process of linguistic selection.

2.4.2 Multiple branching structures

The third example of the simulation in action involves a universal discussed in detail by Hawkins (1994a:§5.2.1) and tested by Kirby (1994a:§4.1.2), involving the orders of noun, adjective and relative clause in NP. If the relative clause is comp-initial, then the noun and the adjective both precede the relative clause. If the relative clause is comp-final, then the noun and the adjective will probably both follow the relative clause, although there are a few exceptions (Hawkins lists Lushei, Dyirbal, Yaqui and Hurrian) in which both precede. In no languages does the relative clause appear between the noun and the adjective as a basic order. Once again, this set of facts seems readily explicable in terms of Early Immediate Constituents: the distance between the first and last of the three MNCCs of the ICs of the NP (N, Adj and Comp) is minimised. The worst cases are where the first MNCC is the first word of the clause and the last MNCC is the last in the clause.

The simulation was tested once assuming relative clauses were comp-initial, and once assuming they were comp-final. In each case there are six competing variants, their relative probabilities of making it into the child's trigger experience being determined by their EIC values (assuming a four-word relative clause). For these first runs, the initial speech community has equal numbers of each variant. The results, consistent with the universals above, are shown in figures 2.9 and 2.10.

[10]Notice that the manner in which this is done is by a kind of gradient ascent. A peak is reached not by the quickest route (directly along one edge of the space), but by following the steepest gradient at each point.

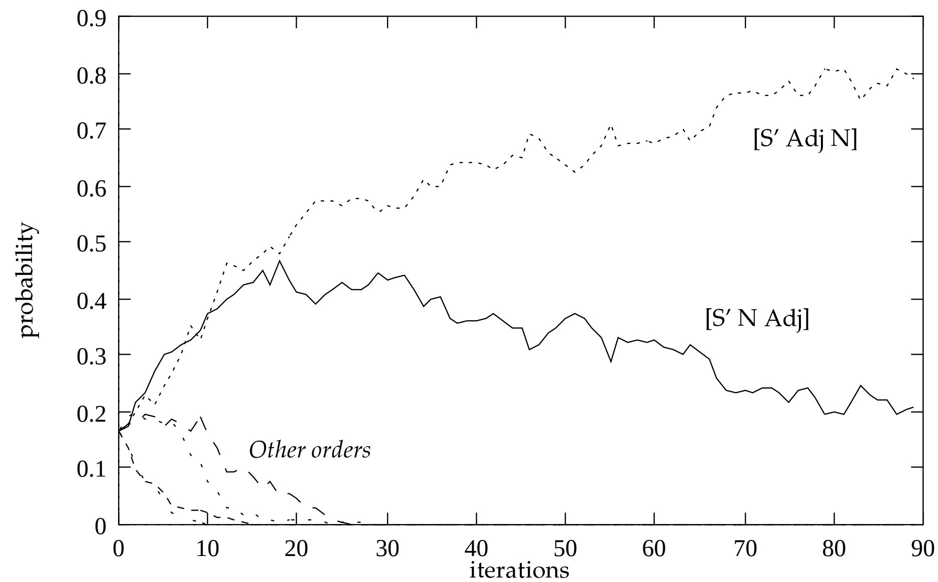


Figure 2.9. A run of the simulation with comp-final relative clause.

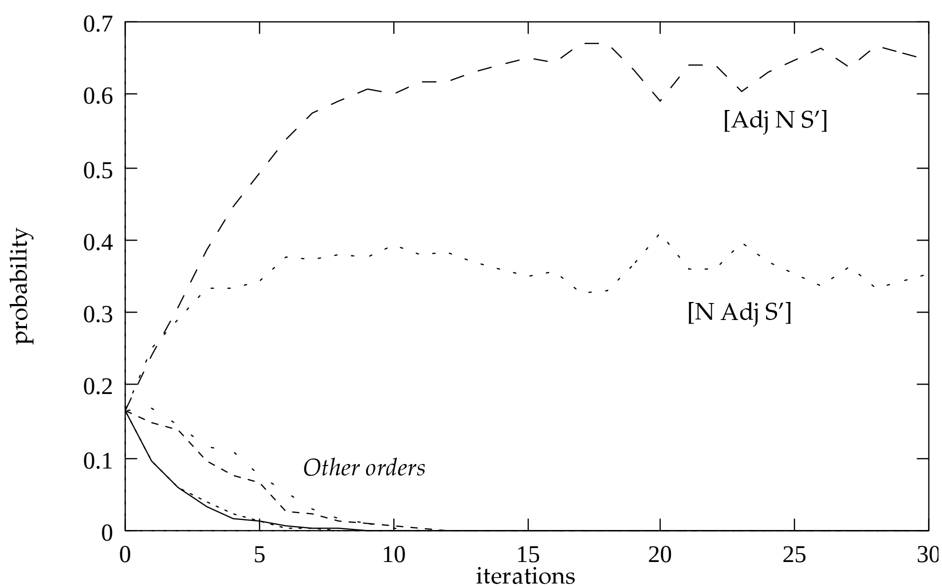


Figure 2.10. A run of the simulation with comp-initial relative clauses.

Notice that in the case of the comp-final relatives the alternative orderings last about twice as long than in the comp-initial case. This is because of an inherent left-right asymmetry in the calculation of the EIC metric. The best non-optimal orderings for comp-final relatives, $_{NP}[N\ Adj\ _{S'}[S\ Comp]]$ and $_{NP}[Adj\ N\ _{S'}[S\ Comp]]$, both have a metric of 0.81 (for a four word relative clause) whereas the orderings $_{NP}[Adj\ _{S'}[Comp\ S]\ N]]$ and $_{NP}[N\ _{S'}[Comp\ S]\ Adj]$ both work out at 0.68. As noted above the exceptions to the relevant universals unsurprisingly involve the sub-optimal orders for comp-final relatives. If the comp-final simulation is re-run with the optimal orders held at zero, the sub-optimal orders eventually "win out" over the worst orderings: $_{NP}[N\ _{S'}[S\ Comp]\ Adj]$ and $_{NP}[Adj\ _{S'}[S\ Comp]\ N]$. This is true even if the original state of the speech community is biased towards these non-occurring orders (figure 2.11). This result suggests that a language that has $_{NP}[N\ _{S'}[S\ Comp]\ Adj]$, say, as its basic order will change its word order given any introduction of variation (except $_{NP}[Adj\ _{S'}[S\ Comp]\ N]$). This means that these worst-possible orders will not be likely to survive very long in any language, and this is reflected in the synchronic universals.[11]

These results also raise the question of what happens when two variants are equivalent in terms of parsing complexity (as are the optimal orders in these examples). The simulation does not converge on a single outright winner in a reasonable time. Instead one order is stable as a minor variant. From this we might predict that wherever there are variant forms of equivalent processing complexity there will always be stable variation. However, this would be a mistake. Labov (1972), for example, discusses a case (the famous Martha's Vineyard study) of a particular sound change in which one variant form clearly wins out over another even though there is no clear processing advantage. Instead the change must be understood in sociolinguistic

[11]This raises some interesting questions about the origin of variation — the other side of the coin as regards a selectionist explanation. These issues are not covered in depth in this thesis, however we can imagine a language contact situation which would introduce a minor variant into a speech community. The important point is that, given multiple competing variants it is possible that the optimum variant may not be available for selection, in which case the "next-best" sub-optimal variant may be selected. Of course, the chances of this happening (and the length of time such a variant survives) will be dependent on its parsing complexity, as shown here.

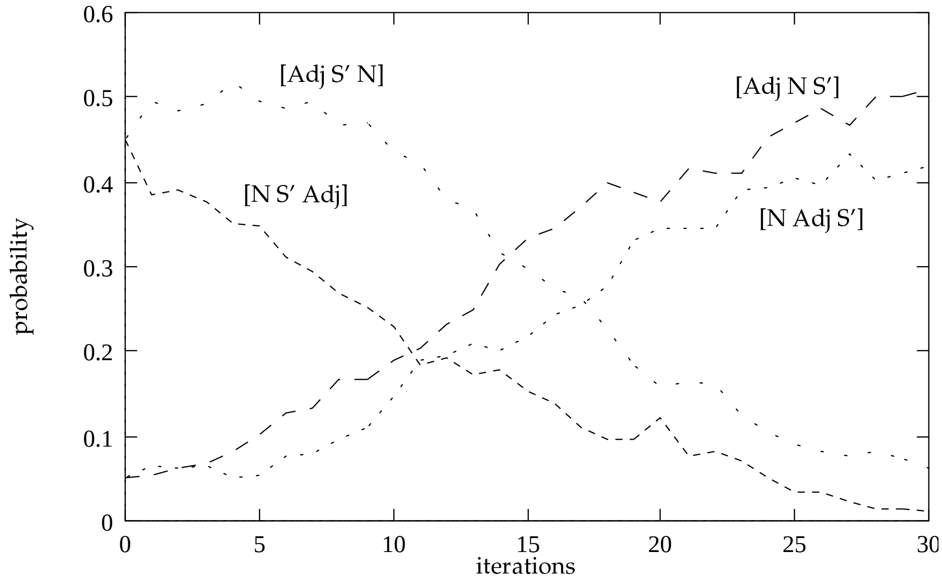


Figure 2.11. Sub-optimal orders in a comp-final language.

terms. Briefly, one form is considered the prestige variant and it is this asymmetry that drives the change (see also McGill 1993, for discussion of this example in terms of selection). Which particular form becomes the prestige variant in this and other such cases is arbitrary with respect to the form itself. So, although sociolinguistic considerations such as these are crucial for understanding change from a microscopic point of view, they do not inform an explanation of universals. We can imagine one of the optimal orders in the simulations above winning out by becoming associated with some sociolinguistic variable, but since the process of association is arbitrary, we can assume that a particular form will be grammaticalised 50% of the time.[12]

[12]One angle for future research might be to see how often this type of selection becomes relevant. In this way it might be possible to predict the frequency of cases where a minor variant survives for an appreciable time. The symmetry of the two optimal variants cross-linguistically will always be preserved, however, as long as sociolinguistic selection has an arbitrary connection to form.

2.4.3 The prepositional noun-modifier hierarchy

The final example in this chapter is somewhat different from the others since it involves pairs of variants whose fitness is independent of each other. The pairs are noun-adjective order, noun-genitive order and noun-relative order within NP. These form a hierarchical universal, the Prepositional Noun-Modifier Hierarchy (Hawkins 1983):

> In prepositional languages, within the noun-phrase, if the noun precedes the adjective, then the noun precedes the genitive. Furthermore, if the noun precedes the genitive, then the noun precedes the relative clause.

$$Prep \rightarrow (NRel > NGen > NAdj) \text{ or...}$$
$$Prep \rightarrow (AdjN > GenN > RelN) \text{ (the contrapositive hierarchy)}$$

This hierarchy predicts that, if a language has structure n in the following list, then it will have all structures less than n.

1. $_{PP}[P\ _{NP}[Adj\ N]]$
2. $_{PP}[P\ _{NP}[NP\ N]]$
3. $_{PP}[P\ _{NP}[S'\ N]]$

Furthermore, according to Hawkins' sample, if a language allows NMod *and* ModN in structure n, then all structures less than n will be allowed but no structures greater than n will (e.g. French: AdjN/NAdj, NGen, NRel or English: AdjN, GenN/NGen, NRel).

How can EIC make sense of these observations? Hawkins (1994a) shows that the EIC metrics of the structures declines down the hierarchy if the lengths of the preposed constituent *increase* down the hierarchy. This is because the distance increases between the MNCC of the first IC of the PP, the preposition, and the MNCC of the last IC, the noun. The simulation takes the length of Adj to be 1 word, Gen to be 2 words, and Rel to be 4 words. The result is shown in figure 2.12. (Notice that the initial situation is set to be at one end of the hierarchy. Kirby (1994a) suggests that this could occur if

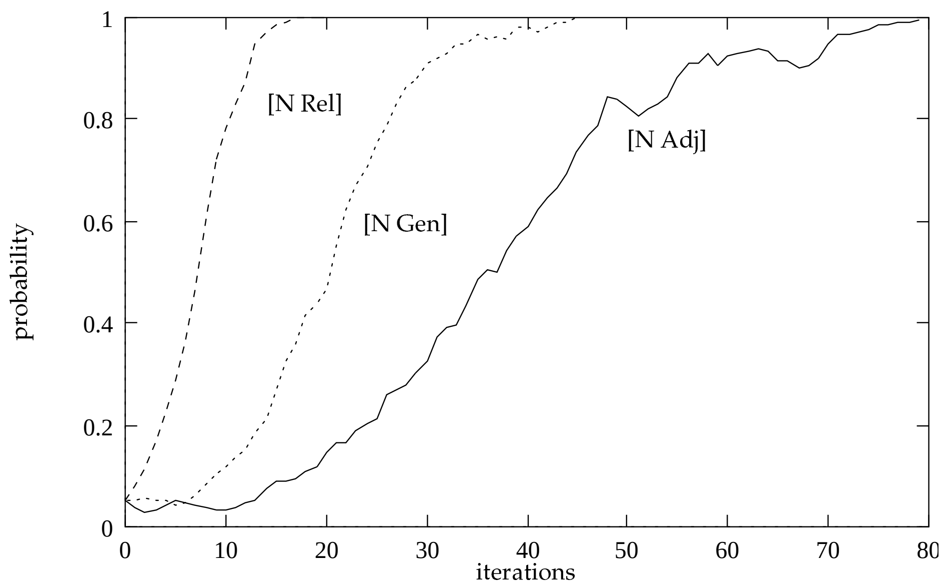


Figure 2.12. Change over time of three independent variant pairs.

a consistently head-final language changed its adposition order. There may be some problems with this, however, which will be discussed further in the next chapter.)

Another way of visualising these same results will show the implicational hierarchy more clearly. Figures 2.13 and 2.14 show the various states of the speech community over the course of the run. The four quadrants of the graph are labelled by language type assuming that the *conventional moment* (the point in time where a speech community is regarded as changing its grammatical conventions) occurs when the probability of a form is greater than 0.5. The quadrants which are not entered by the simulation are GenN&NAdj and RelN&NGen, exactly the types predicted not to occur by the implications underlying the PrNMH: $GenN \rightarrow AdjN$ and $RelN \rightarrow GenN$.

Finally, if a prepositional language has two basic orders for a particular modifier in NP, then it is likely that it is this modifier that is in the process of being preposed. If we arbitrarily section off part of the graph 2.12 around the 0.5 probability line as the area where we might expect free word order for a constituent, then the second typological observation is supported. If the area we choose is between 0.4 and 0.6, say, then after 5 iterations, the speech community has the types AdjN, GenN, RelN/NRel;

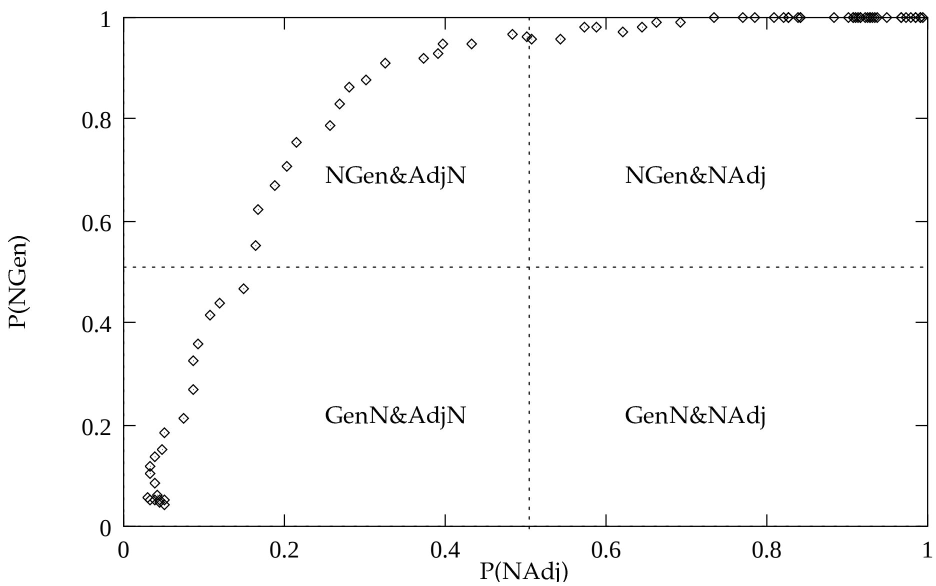


Figure 2.13. Plot of the simulation on NAdj/NGen space.

after 20 iterations, AdjN, GenN/NGen, NRel; after 35, AdjN/NAdj, NGen, NRel.

The simulation results in this section show that the selection model can, in conjunction with his performance metric lend support to Hawkins' Basic Order Prediction, derive the S-shaped logistic curve, and provide a simple explanation for the facts relating to the PrNMH (though see the discussion in chapter 3). Of course, this does not demonstrate that Hawkins' theory is correct; in a sense the argument is a methodological one, demonstrating that viewing language as a complex adaptive system solves the Problem of Linkage. The remainder of this chapter looks at some of the further implications of adopting this position.

2.5 Unifying markedness correlates

As mentioned earlier, Matthew Dryer (e.g. Dryer 1992) uses a method of discovering statistical universals involving counts of *genera* (genetically related language groups of a time-depth no greater than 4000 years) grouped geographically, that is intended to compensate for genetic and areal bias.

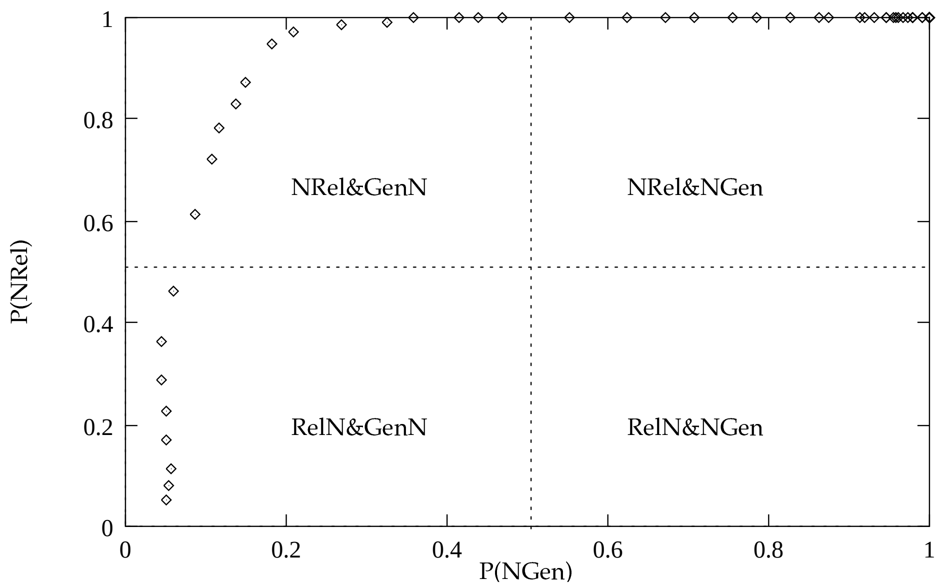


Figure 2.14. Plot of the simulation on NGen/NRel space.

On the basis of this improved method of gathering word order correlations, Dryer argues against the generalisation that *heads* tend to order consistently on one side or other of their dependents. Instead he demonstrates that it is *branching direction* that is relevant:

> "**Branching Direction Theory** (BDT): . . . a pair of elements X and Y will employ the order XY significantly more often among VO languages than among OV languages if and only if X is a nonphrasal category and Y is a phrasal category." (Dryer 1992:89)

Dryer points out that this theory is, in the main, consonant with Hawkins' EIC predictions, which prefer consistently left- or right-branching structures. The main difference is that BDT makes weaker predictions than EIC which includes predictions about left/right asymmetries. These asymmetries should be investigated more closely using Dryer's statistically less biased method.

For our purposes, Dryer's BDT is suggestive of the way in which the adaptive model might be applied to the explanation of why certain criteria for markedness tend to correlate, not only with respect to word-order, but in other domains also.

Given a universal of the type $P \rightarrow Q$ we may say that P is *marked* with respect to Q. This leads us to expect a cluster of linguistic properties associated with markedness to be manifested by P to a greater extent than by Q. Some of these properties as claimed in the literature are listed below:

Structural the more marked value of a grammatical category will be expressed by at least as many morphemes as the less marked category. (Croft 1990:73)

Behavioural (cross-linguistic) if the more marked value occurs in certain language types, then the less marked category will occur in at least those types. (Croft 1990:83)

Frequency (textual) if the more marked value occurs with a certain frequency in a text sample, then the less marked value will occur with at least that frequency. (Croft 1990:85)

Acquisition the more marked value will be acquired later in child language acquisition than the less marked value. (Witkowski & Brown 1983:569)

Language change the more marked value will be added later and lost sooner than the less marked value in language change. (Witkowski & Brown 1983:569)

The structural criterion for markedness is identified by Greenberg (1966:26), following Jakobson's earlier work, as *zero expression*:

> "An important further characteristic of the marked/unmarked opposition . . . I shall refer to . . . as zero expression of the unmarked category.. . . Thus parallel to the example *man* (unmarked), *woman* (marked), we have *author* (unmarked), *authoress* (marked) in which *author* indicates either a writer regardless of sex or specifically a male writer, whereas *authoress* designates only a female writer. In this latter instance the unmarked term *author* has a zero where the marked term *authoress* has an overt affix *-ess*."(Greenberg 1966:26–27)

Notice that Greenberg is essentially defining structural markedness in terms of the *number* of morphemes in an expression. Croft notes that this means that the structural criterion is not a particularly useful one.

> "It is the best-known criterion for markedness in typology. Nevertheless, it is actually of somewhat limited application — for example, we cannot say which of the word orders RelN or NRel is marked on structural criteria — and possibly cannot be applied to phonology. Hence, it is a mistake to identify markedness solely with structural markedness."(Croft 1990:72–73)

This raises the question: can structural markedness be extended to include more than simply the number of morphemes? Here, the inclusion of *complexity* as a markedness criterion in Witkowski & Brown (1983) is the key. If a higher number of morphemes is a reflection of an increase in morphological *complexity*, then perhaps the *configuration* of those morphemes is also a factor in that complexity and hence a candidate for signalling markedness.

I propose, then, that the structural criterion for markedness may be extended to include word order:

Structural markedness (configuration) if the more marked value involves a structure with a certain degree of branching coherence, then the less marked value will involve at least as high a degree of coherence.

Some explanatory remarks are in order here. Branching coherence refers to the extent to which a structure is consistently left- or right-branching, hence the structures ${}_a[{}_b[{}_c[\gamma\ \delta]\ \beta]\ \alpha]$ and ${}_a[\alpha\ {}_b[\beta\ {}_c[\gamma\ \delta]]]$ are maximally coherent whereas ${}_a[\alpha\ {}_b[{}_c[\gamma\ \delta]\ \beta]]$ and ${}_a[{}_b[\beta\ {}_c[\gamma\ \delta]]\ \alpha]$ are minimally so. The word "involve" in this definition is problematic because the markedness of, say, NRel over RelN cannot be judged without examining the context of these structures. In other words, in VO languages NRel is less marked than RelN, but in OV languages the reverse in true. This is an example of *markedness reversal*.[13]

[13]Croft (1990:135) points out that this general phenomenon has been called by various names in the literature such as *local markedness* (Tiersma 1982) and *markedness assimilation* (Andersen 1972).

If branching coherence reflects parsing preference, as Dryer believes and Hawkins' theory predicts, then the adaptive model correctly predicts that the various criteria listed above will correlate. For example, an adpositional phrase within a verb-initial verb phrase may have two orders: $_{VP}[V\ _{PP}[P\ NP]]$ or $_{VP}[V\ _{PP}[NP\ P]]$. The latter of these orders is structurally marked with respect to the former because of its mixed branching — it is also harder to parse by Hawkins' EIC. These two possibilities correspond to the graph in figure 2.6. If the points on the graph correspond to possible human languages,[14] then the **frequency** and **behavioural** criteria apply. Furthermore, if we imagine a language in transition between points on the graph, then the **language change** criterion follows. Finally, although there is no explicit discussion of order of acquisition within the model, we may expect a form which is filtered out of the acquisition/use cycle more often to be successfully acquired later than a form that is not.

2.6 The assumption of speaker altruism

We have seen from the computer simulations in this chapter, that combining Hawkins' performance theory with a theory of linguistic selection goes a lot of the way towards an explanation for word-order universals viewed as phenomena of the third kind. By assuming that the effect of parsing complexity is to influence the transformation of language data into trigger experience (transformation **T3** in figure 2.4) we have a mechanism for solving the problem of linkage. A sensible question to consider at this point is what all this effort has bought us — what does this model add to the explanations in Hawkins 1994a apart from the various goals set out in chapter 1? The main point on which this work differs from Hawkins' is connected with the role of the speaker in explanation. In line with Occam's Razor the selection model so far has not had to call on the speaker to explain the adaptedness of languages since hearer selection is enough. Hawkins, however, implicitly makes use of what I will call *the*

[14] This is a rather crude assumption which needs further justification in this case, as will be argued in the next chapter.

assumption of speaker altruism.

For example, in Hawkins 1994a we find:

> "[Implicational] hierarchies define the sequence in which grammatical variants are selected within each grammatical domain, and the claim is being made that this sequence involves increasing complexity, and that the cut-off points represent a conventionalised response by speakers of each language not to tolerate processing difficulty or inefficiency below that point." (p. 435)

This suggests that the link between processing and competence grammars is the speaker, however, the complexity metrics discussed by Hawkins are measures of *parsing* complexity. If the locus of explanation is the speaker, this suggests that she is responding to the needs of the hearer in her choice of utterance. As Hawkins puts it, "there is, of course, a general benefit for the producer if his or her speech is optimally packaged for the hearer, since communication will then be effective." (p. 426) For this to be the case, however, the speaker must calculate at each choice point in production the *parsing* complexity of the string about to be produced. Now, this may indeed be what is happening — our knowledge of the mechanisms of production is not such that we can know for certain at the moment — but in the light of the explanation put forward, we can afford to be agnostic on this point.

Levelt's (1989) useful review of the experimental evidence relating to production puts forward a *modular* view of the production process that further casts doubt on the assumption of speaker altruism. He breaks the process down into two main stages: *conceptualisation* and *formulation*. The first stage involves the intentional construction of a *preverbal message* which requires information from a discourse model, situational knowledge and so on. This message is passed to the second stage which constructs a *phonetic plan*. Crucially, given the modular approach, the mapping from message to phonetics is non-intentional and does not have access to situational knowledge.

> "Grammatical encoding takes a message as input and delivers a surface structure as output. It is likely that this process is highly automatic and

> non-intentional. A speaker will not, for every message, consider which of various grammatical alternatives would be most effective in reaching some communicative goal." (Levelt 1989:282)

So, even if one of the speaker's communicative goals is to present utterances that are easy to parse, it is not possible that this can affect the choice of grammatical alternatives. To put it another way: to the extent that the preverbal message contains information about the intended order of presentation of a phonetic plan, the choice of a particular order cannot be responsive to the final syntactic form. The conceptualiser therefore cannot make EIC calculations, and the formulator will not be responsive to the needs of the hearer.

However, Levelt's model also includes a monitoring system whereby phonetic plans may be parsed by the speaker and fed back to the conceptualiser. That this self-monitoring is going on is clear from data on speech errors and corrections. Levelt (1989:460–463) gives some examples that suggest this might be a way for speaker altruism to get in by the back door so to speak. For example, when expressing a path through a set of coloured circles in one experiment, a speaker made the following "repair" (from Levelt 1983):

(2.9) We go straight on, or — we enter via red, then go straight on to green.

Here, the speaker makes an error in the ordering of the two clauses which express the sequence of actions to be made in an iconic fashion. This error appears to be caught by the speaker's own parsing mechanisms which signal the need for a repair to the conceptualiser. Another example of word order repair is (from Fay 1980):

(2.10) Why it is — why is it that nobody makes a decent toilet seat?

Again, self-monitoring signals the need for a repair, although in this case the speaker is aware of a syntactic error in the ordering of the subject and copula. Although these repairs seem to offer us a mechanism by which speakers can be responsive to the needs of hearers, it should be noted that all the examples given be Levelt are responses to *errors* rather than hard-to-parse outputs. "Do speakers actually attend simultaneously to all these aspects of their speech? This is most unlikely, and there are data to support the view that ... much production trouble is not noticed by the speaker" (Levelt 1989:463).

Another possibility is that the pressures on language production (i.e. formulation) are simply the same as those on parsing. For such a story to work, speakers would have to prefer to 'build' constituent structure as rapidly as possible. So, a preference for minimal constituent production domains is predicted in parallel with the hearer's preference for minimal recognition domains. The problem with this approach is that the information available to speakers and hearers is radically different, so when producing a verb-final verb phrase, the speaker *already knows* that a VP node can be constructed, whereas the hearer must wait for the MNCC, the verb. Thus this speaker oriented approach fails to predict the structure of languages such as Japanese (Hawkins 1994a:426).

It would therefore seem safer to try to formulate a solution to the problem of linkage that does not assume speaker altruism, and this has been the goal of this chapter. The next chapter returns to the role of the speaker in explaining language universals, though it will be argued not that speakers are altruistic, rather that their preferences are in direct conflict with hearers'.

Chapter 3

Implicational hierarchies, competing motivations

The previous chapter examined a solution to the problem of linkage in the domain of word-order universals, using Hawkins' metric of processing complexity as an example of a partial explanation. This chapter extends the scope of the linguistic selection approach by examining an implicational hierarchy in another domain — accessibility to relativisation.[1] Once again, Hawkins (1994a) provides us with a plausible explanation for the cross-linguistic facts in terms of structural complexity, and this will be the starting point for an investigation of the origins of hierarchies in general.

3.1 Relative clauses and structural complexity

The particular hierarchy which this chapter examines in depth was reported some time ago by Keenan & Comrie (1977) in an important paper. They show that the accessibility of noun phrases to relativisation depends on the grammatical function of the gap or resumptive pronoun within the relative clause according to the hierarchy:

Subject>Direct Object>Indirect Object>Oblique>Genitive>Object of Comparison

[1]The majority of this chapter will appear as Kirby 1996.

This *Relative Clause Accessibility Hierarchy* (AH) constrains possible languages according to the following definitions and constraints:

Subject relative universal "All languages can relativise subjects." (Comrie & Keenan 1979:652) [A strategy that can relativise subjects is a **primary strategy**.]

Accessibility hierarchy constraints

1. "If a language can relativise any position on the AH with a primary strategy, then it can relativise all higher positions with that strategy.
2. For each position on the AH, there are possible languages which can relativise that position with a primary strategy, but cannot relativise any lower position with that strategy." (Comrie & Keenan 1979:653)

Keenan & Hawkins (1987) report results from a psycholinguistic experiment testing native English speakers' 'mastery' of relative clauses down the AH. The experiments were designed to test repetition of RCs that occurred modifying subjects in the matrix clause, so no conclusions can be drawn about: a) other languages, b) RCs modifying matrix objects etc., or c) whether the AH affects production, or perception, or both. These points aside, the mastery of RCs clearly declined down the AH. As Keenan and S. Hawkins point out, this processing difficulty might explain the AH. Other experiments have been carried out that have tested the relative processing difficulty of RCs on the first two positions of the hierarchy (subject and direct object). MacWhinney & Pleh (1988) review a number of studies in comprehension in English children that are consistent with the view that subject relatives are easier to parse than object relatives (though see chapter 4 for further discussion). Furthermore, their own study of Hungarian reveals a similar pattern.

Hawkins' 1994a explanation of this universal relies on these claims that the ease of parsing of relative clause constructions decreases down the hierarchy, and that this leads to the implicational constraints on cross-linguistic distribution. The intuition is that languages somehow select a point on a hierarchy of parsing complexity below which relative clauses will be grammatical, and above which they will be ungrammatical (this approach, then, involves the implicit assumption of speaker altruism). What

Hawkins adds to the work summarised above is an independent theory of structural complexity from which the parsing results can be derived. It is just such a theory that previous attempts to explain the AH (e.g. Kirby 1994c) have lacked.

The theory is related to Early Immediate Constituents in that it defines a measure of tree-complexity associated with a particular node in a constituent that is relative to a particular psycholinguistic operation. In this case this operation is relativisation, rather than constituent recognition. The complexity of relativisation — or rather, processing a relative clause — is proportional to the size of a portion of the tree that is involved in co-indexing the trace, or pronoun, in the clause with its head noun. Hawkins' definitions (pp. 28–31) are as follows:

Structural complexity of relative clause The structural complexity is calculated by counting the nodes in the *relativisation domain*.

Relativisation domain The relativisation domain consists of that subset of nodes within the NP dominating the RC that structurally integrate the trace or pronoun.

Structural integration of a node X in C The set of nodes which structurally integrate X in C are:

- all nodes dominating X within C (including C itself)
- all sisters of X[2]
- all sisters of the nodes dominating X within C

The intuition captured by this definition is that relating the head noun with a trace (or pronoun) becomes more complex the more the trace (or pronoun) is embedded within the subordinate clause.

Hawkins demonstrates this metric using tree structures that rely on traditional notions of constituency, but the complexity rankings seem to remain the same if they are calculated using other syntactic analyses. Consider the structures in figures 3.1

[2]In fact, some sisters may be excluded from the calculation if the language has flatter configurational structure. In this case, morphological case contributes to the calculation of structural complexity. For example, in languages without VPs, nominative marked NPs may be included as sisters of an accusative, but not vice versa. See Hawkins 1994a, 27–28 for discussion.

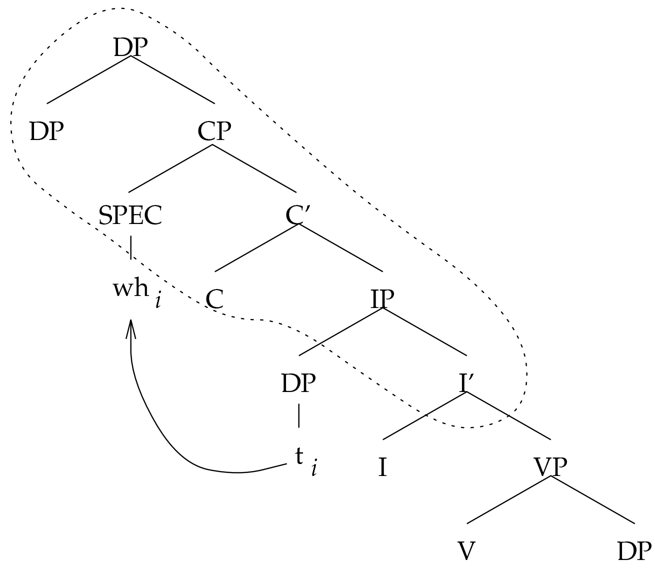


Figure 3.1. Subject relative.

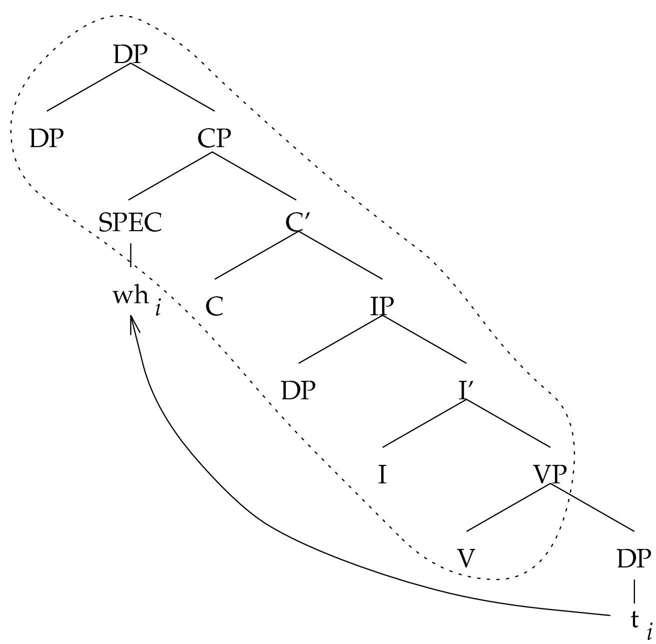


Figure 3.2. Object relative.

and 3.2, which are standard treatments of relative clauses within the Principles and Parameters tradition. The first tree is a structure where the subject DP in [Spec,IP] has moved to [Spec,CP]. This, then, is the structure of a subject relative. The nodes that are involved in the calculation of complexity are circled. The second tree is the equivalent for an object relative — in this case, it is obvious that the RC-complexity is higher. Similar arguments can be made for the relative ranking of other positions on the AH (Hawkins 1994a:39–41).

This account is successful inasmuch as it predicts the relative ranking of relative clauses in a hierarchy of parsing complexity, and uses concepts — such as structural domains — which can be generalised to other domains (e.g. word order and extraction). However, the theory as it stands does not answer the problem of linkage; exactly how do the structural complexity facts end up being expressed cross-linguistically? The next section attempts to answer this question in the same way as in the previous chapter, and in doing so shows that structural complexity *cannot on its own* give rise to hierarchy.

3.2 Extending the computational model

The simulation approach used here is almost identical to that of the previous chapter; the only real change is in the structure of the Arena of Use. So far the simulations have been used to examine the time course of changes in a speech community. These have all resulted in a reduction of variation over time, leading eventually to a homogenisation of the community — this has been referred to as grammaticalisation of one of a set of variant options. In a sense we have seen the operation of the invisible hand as an emergent property of the simulations. However, recall that universals have been characterised as *higher order* emergent properties, and as such we would prefer to see them emerge as stable end states of the simulations. This is not possible if the only end states are homogeneous (i.e. if the simulations always converge on a single type).

The element in the simulations that gives rise to homogenisation seems to be the Arena of Use. All speakers input to an unstructured Arena, and all hearer/acquirers take input from random points in the Arena. Thus any differences in a population of speakers will be 'averaged-out' in the next iteration of the simulation. For there to be a stable end state with multiple types it must be possible for structure to emerge and be sustained in the population. The simulation described below achieves this by arranging speaker's spatially and dividing the Arena of Use into many overlapping, localised Arenas.

3.2.1 A new simulation

The simulations discussed here examine only the first two positions on the accessibility hierarchy — subject and direct object. Discussion in later sections shows how these results are easily extended to the rest of the hierarchy, and provide an explanation for the **subject relative universal** that we will ignore for the moment. The relevant components of the simulation are:

Utterances The E-domain objects of the simulation. Either S, O, S' or O', corresponding to utterances with subject relatives, with object relatives, without subject relatives, and without object relatives.

Arena of Use A two-dimensional toroidal[3] space of utterances arranged such that an utterance at coordinates (x, y) was uttered by a speaker at (x, y).

Grammars These are the I-domain objects. They are either SO, $S'O$, SO', $S'O'$ corresponding to the four possible language types.

Speakers A speech community is made up of a two-dimensional toroidal space of speakers each of which consists of a grammar.

Acquirers These are speakers without grammars. They take input from nearby coordinates in the Arena (as described below).

[3]So a cell in the space has neighbours above, below, to the left and to the right. A cell on the bottom edge of the space has a neighbour at the top and vice versa, and a cell on the left edge has a neighbour on the right edge and vice versa. This geometry is chosen mainly because it is easily implemented, not having 'edges'.

The dynamic processes involved are:

Production Speakers add utterances at random to the point in the Arena at the same coordinates as themselves in line with their grammars.

Parsing/Acquisition Acquirers become speakers in the following way:

1. The neighbouring speakers' coordinates are found, where an acquirer has 4 neighbours: one above, one below, one to the left, and one to the right of its position.
2. All the utterances from the arena at the neighbours' positions and at the position of the acquirer are pooled together and a random subset is taken to form the linguistic data input to acquisition.
3. This data is filtered to form a trigger. This process involves measuring the relative distribution of variants in the data, and then choosing from those variants in such a way as to reflect its distribution and its relative structural complexity.
4. The trigger is then mapped directly onto the acquirer's grammar.

As with the simulations in the previous chapter, a run involves each speaker producing some number of utterances, and then each acquirer parsing/acquiring on the basis of the arena (although with this simulation the relevant data will be that produced 'nearby' the acquirer). After acquisition, the old speakers and Arena are discarded and replaced by the acquirers and the process is repeated.

3.2.2 Testing the explanation

If the explanation of the accessibility hierarchy based on a parallel hierarchy of structural complexity is correct, we should be able to run the simulation and see the implicational universal $O \rightarrow S$ emerge. To test this, the simulation was set up using the following equations to produce the trigger:

$$p(S) = \frac{w_S n_S}{w_S n_S + (1 - w_S) n_{S'}}$$

$$p(O) = \frac{w_O n_O}{w_O n_O + (1 - w_O) n_{O'}}$$

where

$$w_O < w_S < 0.5$$

This means that both object relatives and subject relatives are dispreferred in terms of parsing to non-relativised alternatives (we will come to what those alternatives might be later), and that object relatives are harder to parse than subject relatives. The actual values seem to affect only the rate at which the simulation converges to a stable end point and the sensitivity of the simulation to initial conditions. The values used for the results shown here were $w_S = 0.4$ and $w_O = 0.3$. The initial speech community was always set to a random spread of all four possible language types.

The first feature of the simulation results, which is largely independent of the initial conditions, is that large groups of similar individuals – language communities – quickly form. This is a similar result to one of Jules Levin (reported in Keller 1994, 100). Levin's simulation is similar to this one in many respects, but it does not model the influence of selection in parsing or production (transformations **T1** and **T3** in figure 2.4). In other words, it assumes that the language that an individual will acquire is simply the one that most of that individual's neighbours has. Keller (1994:99) calls this 'Humboldt's Maxim':

> "Talk in a way in which you believe the other would talk if he or she would talk in your place. My thesis is that this maxim — a slightly modified version of Humboldt's own formulation of it — produces homogeneity if the starting point is heterogeneous and stasis if the starting point is homogeneous."

Indeed, this is what happens with Levin's simulation. Starting with a random patterning of two types, the simulation finally settles down with the types clustering together in large groups. (Homogeneity here does not mean complete lack of variety, there are still two types, rather variation has decreased spatially.)

The result of a typical run of the simulation described here is shown in figure 3.3.

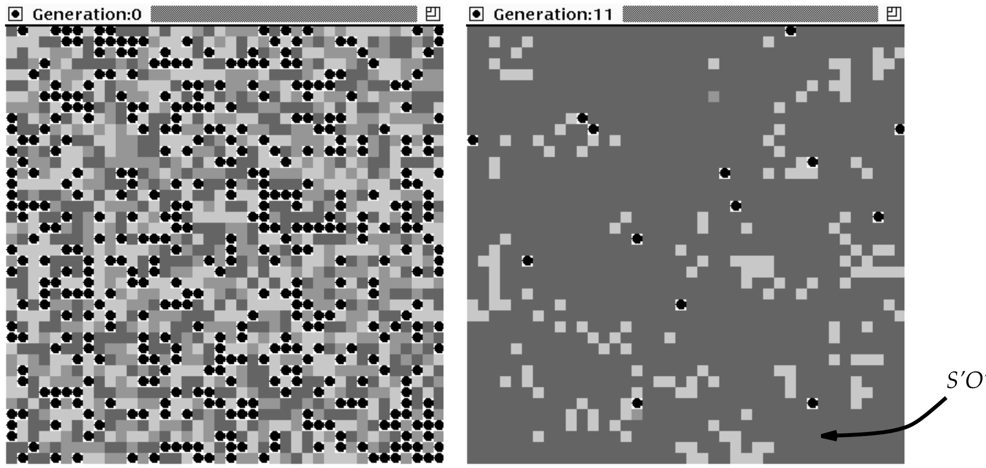


Figure 3.3. The initial (random) and eleventh ($S'O'$ only) generations of a simulation run.

Each small square on the figure is a speaker in the simulation, and the shading for the squares indicates one of the four possible language types. The expected result, if a gradient hierarchy of complexity can explain the accessibility hierarchy, is that the end result should show the types SO, SO' and $S'O'$ (recall that the subject relative universal is ignored for the moment). The only type that should not survive is $S'O$. For clarity, speakers with grammars of this type are indicated by black circles in the diagram. The problem with the results in 3.3 (and with all such runs of the simulation) is that the community converges on only one type: $S'O'$. This clearly poses a serious problem for the complexity hierarchy explanation.

3.3 Competing motivations

The solution to this problem involves a 'competing motivations' explanation (e.g. Hall 1992; DuBois 1987; Givón 1979). These are explanations that rely on functional pressures in *conflict*. Newmeyer (1994a) examines several different types of these explanations and argues that *some* attempts by functionalists to build these sorts of motivations directly into their theories of synchronic grammatical phenomena render both their descriptions and their explanations inadequate. These criticisms will not apply to the approach taken in this paper since the functional pressures in question are not assumed to be encoded in grammars. Instead, the I-language domain is taken to be autonomous from the environment;[4] however, as the model described in the previous section makes clear, this does not preclude the possibility that functional pressures can influence the possible states a grammar can take.

3.3.1 Types of complexity

The influence on parsing of structural complexity is *one* functional pressure that affects relative clauses. Because it affects parsing, it is part of what I will call *p-complexity*. The details of a full definition of p-complexity will involve many different aspects, but the influence of it within the selection model is simple:

p-complexity In comprehension, the selection of competing variants (i.e. variant forms that are synonymous, or functionally undifferentiated) will depend on their relative parsing complexity. So, the more difficult some variant is to parse, the more likely it will fail to be included in the set of trigger experiences of the child.

Some of the other factors that influence p-complexity are, for example, redundancy of information (§3.5.1), and configurational markedness (§2.5). Another type of complexity that will influence the selection model is morphological or *m-complexity*:

m-complexity In production, the selection of variants will depend on their relative morphological complexity. So, given two competing ways in which to produce some message, the speaker will be more likely to produce the one that is less morphologically complex.

Traditional structural markedness, where a marked form has more morphemes (see, e.g. Croft 1990, 73, and the discussion in the previous chapter §2.5), is clearly related

[4] However, see chapter 5 for discussion of a mechanism through which features of the environment *can* become encoded in an autonomous grammar.

to m-complexity. However, precisely how this affects production is not clear: is the relevant measure the number of morphemes, or the number of morphs? Do all morphemes carry equal m-complexity, or are morphemes that are involved in agreement (ϕ-features) more complex to produce than others (such as definiteness markers)? We shall return to this question later, but since we will typically be looking at the relative ranking of variants with regard to m-complexity, it is not fatal to avoid specifying the details of its definition, here.

This is a competing motivations explanation, since it claims that the pressures that these factors bring to bear on the selection of relative clauses are opposed. Consider the following Malagasy examples (from Keenan 1972b):

(3.1) ny vehivavy izay nividy ny vary ho an' ny ankizy
the woman REL bought the rice for the children
'the woman who bought the rice for the children'

(3.2) a. * ny vary izay nividy ho an' ny ankizy ny vehivavy
the rice REL bought for the children the woman
'the rice which the woman bought for the children'

b. ny vary izay novidin' ny vehivavy ho an' ny ankizy
the rice REL bought+PASS the woman for the children
'the rice which the woman bought for the children'

(3.3) a. * ny ankizy izay nividy ny vary (ho an) ny vehivavy
the children REL bought the rice (for) the woman
'the children who the woman bought the rice for'

b. ny ankizy izay nividianan' ny vehivavy ny vary
the children REL bought+CIRC the woman the rice
'the children who the woman bought the rice for'

(3.1) is an example of a subject relative in Malagasy. (3.2a) shows that object relativisation in Malagasy is ungrammatical. This raises the question of how speakers get round the problem of presenting the message in (3.2a) without using the ungrammatical relative. The solution in Malagasy is to promote the object to subject using a passive and then relativising on the derived subject (3.2b). This structure is morphologically marked with respect to the non-passivised equivalent since it involves extra passive morphology on the verb, hence it has a higher m-complexity. Similarly, Malagasy oblique relatives (3.3a) are ungrammatical (as we should expect from the AH). Instead, speakers can use another promotion-to-subject construction (3.3b). Here, a "circumstantial" affix is attached to the verb that promotes the oblique object to subject. Again, this clearly involves an increase in m-complexity.

Here, then, is a case where avoidance of some relative causes an *increase* in m-complexity, but a *decrease* in p-complexity.[5] Thus, the two complexity motivations are in competition.

3.3.2 Testing the competing motivations

In order to test what effect m-complexity has on the simulation, the way in which I-language is mapped onto utterances (the transformation **T1** in figure 2.4) needs to be adjusted. It is too simplistic to say that speakers produce utterances in line with their I-language states; instead, the probability of producing morphologically simpler forms should be weighted higher than the higher m-complexity variants. To do this, a variable w_R is introduced that represents the speaker preference of S and O over the higher m-complexity non-relative variants S' and O'. The parameters of the simulation are therefore:

[5] The relative clauses are subject relatives, and thus have smaller structural domains. Hawkins (1994b:31) explicitly states that the calculation of structural complexity should relate to the position of the co-indexed element inside the clause "in its original (d-structure) position" in an attempt to provide a unified account of promotion-type relatives such as (3.2b) and non-promoted relatives. However, there are reasons why we should be wary of this approach and, at least as a first approximation, use a definition that refers to the surface position. One of the results of Keenan & Hawkins (1987) work is that when errors are made repeating relatives, then the errors tend to be towards relatives on higher positions on the hierarchy. The majority of errors made repeating relativised direct objects were RCs on the subject of a passive; the majority of errors made repeating relativised subjects of passives, however, were RCs on direct objects. A possible explanation is that the former case is a response to p-complexity (the RC was mis-parsed), whereas the latter is a response to m-complexity (a simpler paraphrase is produced).

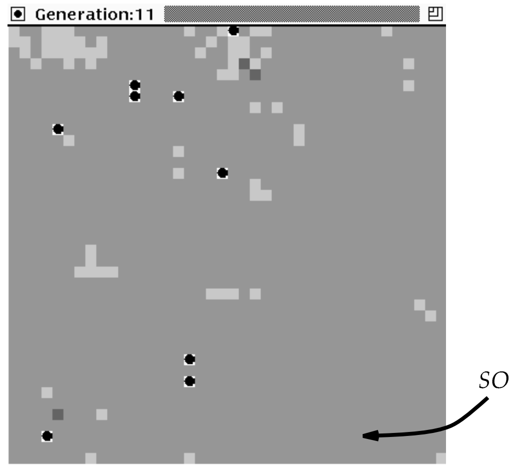


Figure 3.4. The eleventh generation of a simulation run showing SO only.

variable	values	interpretation (inverse of)
w_R	$w_R > .5$	m-complexity of RC variants
w_S	$w_S < .5$	p-complexity of subject RC
w_O	$w_O < w_S < .5$	p-complexity of object RC

Depending on the initial conditions, one of two results emerges depending on the relative magnitude of m-complexity and p-complexity. If m-complexity is high, then the end result is languages of type $S'O'$ only (as in the previous simulation), whereas if p-complexity is high, the end result is languages of type SO only (see figure 3.4). Obviously, with neither starting condition does the hierarchy emerge.

Although this result seems to suggest that the competing motivations hypothesis has failed, this in fact depends on the values of the variables in the table above. These variables are set to certain values at the start of the simulation and remain the same for all points in the simulation space and over time. However, it is not plausible to say that the relative magnitude of m-complexity and p-complexity will be invariant for languages. To see why, compare the Malagasy examples with some Malay examples, also from Keenan 1972b:

(3.4) Ali bunoh ayam itu dengem pisau
Ali killed chicken the with knife
'Ali killed the chicken with the knife'

(3.5) a. * pisau yang Ali bunoh ayam itu dengem
knife REL Ali killed chicken the with
'the knife that Ali killed the chicken with'

b. pisau yang Ali gunaka untok membuno ayam itu
'the knife that Ali used to kill the chicken'[6]

Malay is unable to relativise on obliques (3.4-3.5a), however there is no way in which to promote the oblique to subject as in Malagasy (3.2b). When Keenan's informants were asked to produce an equivalent to the English oblique relative, they gave a paraphrase such as (3.5b).

As well as paraphrase and promotion, circumlocution is another strategy for avoiding relatives. Consider variants to the English (3.7a) and (3.6a).

(3.6) a. I watch the batsman who England selected.

b. I watch the batsman who was selected by England.

(3.7) a. I watch the team which Hick plays cricket for.

b. *I watch the team which was played cricket for by Hick.

c. I watch this team — Hick plays cricket for them.

(3.6b) is the promoted variant of (3.6a), but the passive is not available to promote the oblique and reduce p-complexity (3.7b). Another option in this case is to use something like (3.7c) which does not have a relative at all.

The point of these examples is to show that the relative m-complexity of relative clauses and their non-relative variants really depends on a variety of factors connected with other systems in the language in question. In certain languages like Malagasy, there is a well-developed voice system that enables promotion to subject. Malay,

[6]Keenan does not provide a gloss with this example.

on the other hand, has a less well developed system, and cannot promote obliques. English can promote some NPs, but the passive involves higher m-complexity (is morphologically more marked) than the passive in Malagasy. To sum up, the relative magnitude of m- and p-complexity is not universally fixed, rather it is affected by the systems made available by the rest of the language and may vary over time.

To model this, the simulation is adjusted so that every few iterations the relative magnitude of m- and p-complexity is adjusted for a random language type. This involves introducing another parameter that expresses the probability of a change occurring each iteration, but the value of this parameter does not seem to be too critical. The result of this seemingly small change in the simulation is profound. Instead of settling down to a static end state with only one predominant type like the other simulation runs, the state of the simulation 'world' is constantly changing. Large groups form, as in Levin's simulation, and in my previous simulations, but at the boundaries of these groups something akin to borrowing occurs, and language types move across space, and change prominence over time. A few of the generations in a typical run of the simulation are shown in figure 3.5. The most important feature of these results is that all language types are well represented *except for* $S'O$. (This is the type marked as black circles.) $S'O$ takes up about one quarter of the initial space, by generation 10, however, there is almost none of the type displayed. Over a long run, the other three types (indicated for the final generation) share the space roughly between themselves.

The implicational universal has emerged.

To summarise, the results from the three simulation experiments are:

p-complexity only: static end state – $S'O'$

p- and m-complexity, fixed: static end state – either $S'O'$ or SO

p- and m-complexity, variable: dynamic state – $S'O'$, SO' and SO

These results lend strong support to a competing motivations analysis within a selection model where the magnitude of the selection pressures is variable. The next

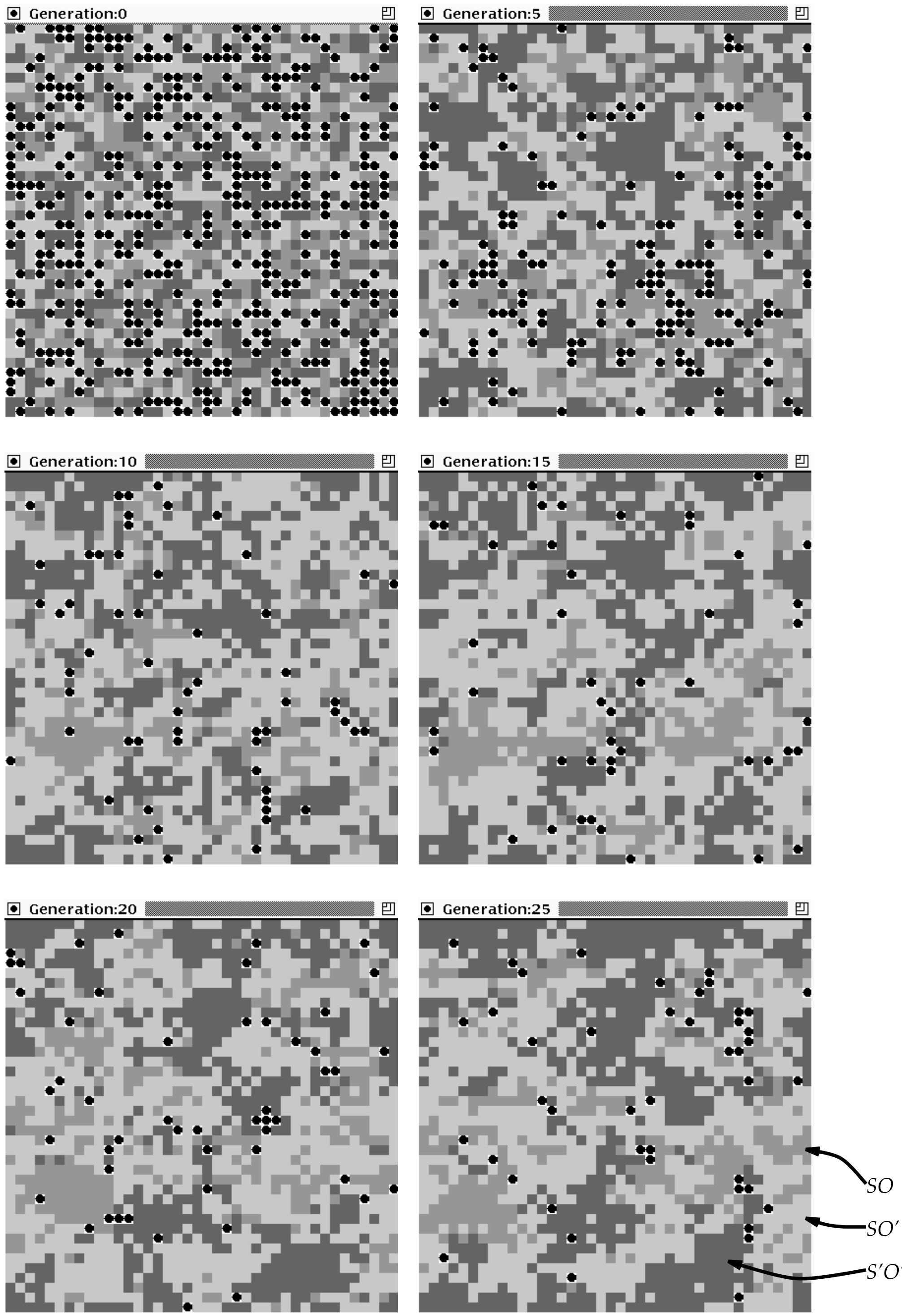


Figure 3.5. An example run of the simulation with shifting complexities. Note that number of the $S'O$ type (here in black) is reduced rapidly from the initial condition. (Proportion of $S'O$ is 27% at generation 0, and 3% at generation 25.)

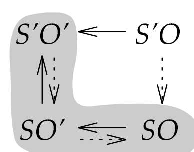
section discusses how this result can be generalised to other positions on the AH, and gives an explanation for the subject relative universal.

3.4 Dynamic typology

In order to understand what the simulation is doing, we need a theory of how dynamic processes give rise to universal constraints. In other words, if we understand what types of changes are likely to occur when the simulation is in one state, then is there a way to calculate what universals will emerge? Borrowing from Greenberg (1978) we will use *type graphs* in order to answer this question.

A type graph is a graph whose nodes are states in a language typology, and whose arcs are possible transitions between those states. So, for the example discussed above, there will be four nodes in the type graph: $S'O'$, $S'O$, SO' and SO. As we have seen, which transitions between these states are possible depends on the relative magnitude of m- and p-complexity. This is represented by two different types of arc: solid ones for when p-complexity considerations are paramount, and dotted ones for when m-complexity outweighs p-complexity:

(3.8)



If we follow the transitions on this graph we can see what happens to a language in the simulation given a particular initial state. So, if a language relativises on subjects and objects, and the m-complexity of RC variants is low, then the next state of the language will be subject-only relativisation, and then neither subject nor object relativisation.[7] Considering only the solid arcs on the graph, then the situation is equivalent to the

[7]This graph only shows what will happen all things being equal – in other words, if there is sufficient random variation in the environment to allow speakers and hearers to freely select variant forms. The simulation described in the last section does *not* make this assumption, however, since variation is drawn from other languages which are also following paths through the type graph.

first run of the simulation where m-complexity was not considered. It is clear that the inevitable end state will be $S'O'$ since once a language is in this state, then it cannot escape. This is termed a *sink* by Greenberg (1978:68). Similarly, if only the dotted arcs are considered, then SO is a sink. This explains why the second simulation run always ended up at one of these two end states depending on the initial conditions.

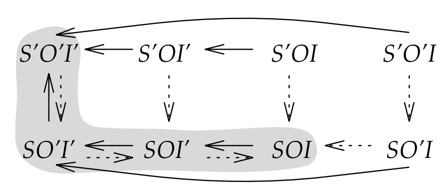
If both types of arcs are considered, then the implicational universal emerges: languages end up in the shaded region of the graph. An informal definition of areas of type graphs that corresponds to universals is given below:

> The language types that are predicted to occur are the set of nodes that belong to strongly connected sub-graphs whose members are only connected to other members of the sub-graph.

A node a is 'connected' to b if there is an arc from a to b, or if there is an arc from a to c and c is connected to b. A graph is 'strongly connected' if for every node a and every node b in the graph a is connected to b (and vice versa). So, in (3.8) all the nodes in the shaded region are connected to each other, but once languages are in this region they cannot escape from it.

The graph can be extended to other positions on the hierarchy. So, for example, (3.9) is the graph for the first three positions on the AH: subject, direct object and indirect object. Again, the universal that is predicted by the definition above is shaded:

(3.9)



The shaded regions in the graphs above are indeed what the accessibility hierarchy predicts.

A problem with this result is that it does not correspond to what is found in reality. This is because of the separate subject relative universal which states that all languages relativise on subjects. This is a case where the type graph theory can be used to look for a possible explanation. The smallest change that can be made to the graph above to bring it in line with the observed universal is to remove the solid arc leading from $SO'I'$ to $S'O'I'$ (i.e. remove the hearer-driven change that makes subject relatives ungrammatical):

(3.10)

$$\begin{tikzcd}
S'O'I' \arrow[d, dotted] & S'OI' \arrow[l] \arrow[d, dotted] & S'OI \arrow[l] \arrow[d, dotted] & S'O'I \arrow[lll, bend right] \arrow[d, dotted] \\
SO'I' \arrow[r, dotted, shift right] & SOI' \arrow[l, shift right] \arrow[r, dotted, shift right] & SOI \arrow[l, shift right] & SO'I \arrow[l, dotted] \arrow[lll, bend left]
\end{tikzcd}$$

In fact, it seems that this might indeed be the correct modification to the previous explanation. Recall that languages typically provide a number of possible ways of 'avoiding' a particular relative clause construction. One of the least morphologically complex of these strategies is the promotion-to-subject strategy exemplified by the Malagasy examples (3.4–3.5). This strategy is *not* available to avoid subject relatives, however, and even if the language allowed demotion this would not be a viable option since it would increase the p-complexity of the relative clause. So, this calls into question an idealisation in the design of the simulation: namely, that relative m- and p-complexity shifts randomly. If promotion is unavailable for subjects, then the average relative m-complexity of constructions that avoid subject relativisation will be higher than for other positions. Selection by the speaker – in terms of m-complexity – will thus be more likely for this position.

3.5 Case coding and complexity

So far only *primary* relativisation strategies have been considered. These are strategies for relativisation that are used for subjects according to Keenan and Comrie's definition. However, languages often make use of different strategies for relativisation on lower positions on the hierarchy. It turns out that the competing motivations approach makes some interesting predictions for the distribution of these strategies.

3.5.1 A strategy taxonomy

Two broad types of relativisation strategy are examined in early work:

The case coding taxonomy: (adapted from Comrie & Keenan 1979 and Keenan & Comrie 1977) A strategy for relativisation is case-coding (or [+case]) if a nominal element is present in the restricting clause which case marks the relativised NP at least as explicitly as is normally done in simple sentences.

An example of a [–case] strategy in Arabic relativisation is given by Keenan & Comrie (1979:333):

(3.11)	al-	rrajul	ya'raf	al-	sayeda	allati	nayma
	the	man	knows	the	woman	REL	sleeps

'The man knows the woman who is sleeping'

Here the relative marker does not code for the case of the NP in the subordinate clause being relativised, and there is no extra nominal element with the clause that marks its case. Object relativisation in Arabic is [+case], however (Keenan & Comrie 1979:333):

(3.12)	al-	walad	ya'raf	al-	rajul	allathi	darabat	-hu	al-	sayeda
	the	boy	knows	the	man	that	hit	him	the	woman

'The boy knows the man whom the woman hit'

In this example, the case is coded by the resumptive pronoun *-hu* within the restrictive clause. Another example of a [+case] strategy is given by standard written English direct object relativisation:

(3.13) The boy knows the man whom we saw.

Here, the relative pronoun marks the relativised NP as a direct object. Notice that the commonly used relative markers (*who, which, that*) occurring in subject and direct object relativisation can all be used for both those positions and are thus [–case], since they do not explicitly code the case of the relativised NP.

In these examples, and universally, [+case] strategies occur lower on the AH than [–case] strategies. This is predicted by the theory outlined in this chapter if we include a notion of information content in the definition of p-complexity. When defining the p-complexity of RCs it was argued that complexity must be relative to a particular psycholinguistic operation – namely the association of the trace, or resumptive pronoun, with the head noun. The complexity of this association task may be ameliorated by providing (typically redundant) information relating to the grammatical function of the embedded element. Hawkins (1994a:45–46) supports a similar analysis: the 'conservation of logical structure' hypothesis of Keenan (1972a). This states that resumptive pronouns make the correspondence between surface structures and logical-semantic structures of relative clauses more transparent, and therefore make processing easier. However, this analysis only covers resumptive pronouns, whereas a treatment in terms of redundancy of information covers the full range of possible [+case] strategies.

The two types of strategy differ with respect to both m- and p-complexity:

[+case] High relative m-complexity (extra nominal element increases morphological markedness), low relative p-complexity.

[–case] Low relative m-complexity, high relative p-complexity.

At first blush, this seems to make no predictions about the distribution of strategies. Again, m- and p-complexity are in conflict. However, the relative markedness of the two strategies changes down the accessibility hierarchy:

Change in relative m-complexity: The typical m-complexity of an RC high on the hierarchy will be lower than that of one low on the hierarchy, therefore any increase of m-complexity will be more marked high on the hierarchy.

Change in relative p-complexity: The low positions on the hierarchy have higher p-complexity, so it is less likely that a form that increases p-complexity further will survive to the trigger on these positions.[8]

[8] Notice that the asymmetry between speaker and hearer selection here is explicable given that speakers make selection 'choices' by comparing the two variants directly, whereas hearers/acquirers do not have direct access to a comparison of the two forms at the point of selection.

It is apparent that case-coding represents a trade-off between an increase in m-complexity and a decrease in p-complexity. For positions low on the hierarchy the balance is in favour of selection in terms of p-complexity (hearer selection) giving [+case] strategies, whereas positions high on the hierarchy favour selection in terms of m-complexity (speaker selection) giving [–case] strategies.

3.5.2 Beyond [+/–case]

Tallerman (1990) revises the definition of [+case] to include examples where the relativised NP is marked without an explicit nominal element. The motivation for this is to analyse examples of consonantal mutation in Welsh – which disambiguate the function of the relativised NP – as [+case]. The new definition also includes strategies that explicitly mark the grammatical function of the relativised NP by word order (e.g. English):

Case coding strategies: (Adapted from Tallerman 1990, 293) A strategy for relativisation is case-coding or [+case] if it explicitly signals the grammatical function of the relativised NP. (Not necessarily with a nominal element.)

In fact, this means that most languages use solely [+case] strategies, in Tallerman's sense, unless word order produces ambiguous relative clauses. Welsh provides examples where there are both [+case] and [–case] strategies, since the basic word order is VSO (Tallerman 1990:296).

(3.14) y bachgen a welodd *t* y ci *t*
the boy COMP saw-3SG the dog
'the boy who saw the dog' or
'the boy who the dog saw'

In this example, the *t*s mark the possible positions for the trace, yielding the two possible readings respectively. This is [–case] relativisation. As mentioned above, Welsh consonantal mutation provides a [+case] strategy (Tallerman 1990:300):

(3.15) y bachgen a welodd *t* **g**i
the boy COMP saw-3SG dog(+MUT)
'the boy who saw a dog'

(3.16) y bachgen a welodd **c**i *t*
the boy COMP saw-3SG dog(-MUT)
'the boy who a dog saw'

Put simply, there is a morphophonemic set of changes in Welsh known as soft mutation which occurs on some segments in certain environments, including directly following a noun phrase. Wh-traces are included in the set of triggering environments, hence the mutation of the initial segment in *ci* above.

An interesting feature of Tallerman's definition of [+case] is that it allows us to go beyond the simple case-coding strategies with opposition between speaker and hearer and look in more detail at the interaction of m-complexity and cross-linguistic distribution. Firstly, a further definition:

Zero-morpheme strategy: A strategy that is case-coding (in Tallerman's sense) but uses no extra morphemes ('nominal elements') for case-coding is a zero-morpheme strategy.

Hence, Welsh soft mutation is a zero-morpheme strategy. Since zero-morpheme strategies are case-coding, with low relative p-complexity, but *without* the concomitant increase in relative m-complexity, we can predict that zero-morpheme strategies will be used as high on the accessibility hierarchy as they can be.[9] This is indeed true in the Welsh case. If the so-called word order strategies in the sample of Maxwell 1979 are taken into account, then this is further support for this prediction since they are all primary strategies.

We can extend the prediction about zero-morpheme strategies by formulating a hierarchy of strategies that is ranked in terms of m-complexity:

[9] This will generally mean that they will be used for subject relativisation (i.e. they will be primary strategies), however it is conceivable that a zero-morpheme strategy may be constrained in other ways so that it cannot be freely selected for on every position on the hierarchy (see also chapter 4).

Strategy hierarchy: [+case] strategies may be ordered with respect to the typical relative m-complexity of case-coding, such that a complex or 'weighty' strategy occurs low on the hierarchy:

Zero-morph > Case-coding Relative Pronoun >? Anaphoric Pronoun[10]
(> Clitic Doubling etc.)

The lower the strategy is on this hierarchy, the lower on the accessibility hierarchy that strategy will occur cross-linguistically.

This hierarchy is rather speculative since there has been no typological research that categorises strategies to this level of detail. The study of Maxwell (1979) refines the Keenan/Comrie sample by categorising strategies as word-order, relative-pronoun and anaphoric-pronoun, among others. Maxwell's categorisation is obviously not motivated by morphological complexity and we must be cautious of any support that his work provides. However, it is interesting to note that the distribution of anaphoric pronoun strategies in the sample is skewed significantly lower on the accessibility hierarchy than that of the relative pronoun strategies.[11]

Even within one language, we can find support for the strategy hierarchy. Looking again at Welsh, Tallerman (1990:313) notes that a pronominal strategy can be used for some direct objects, some non-direct objects and genitives. A clitic doubling strategy, however, is only available for some non-direct objects and genitives. This distribution is expected since the clitic doubling strategy (3.18) has a higher m-complexity than simple retention of an anaphoric pronoun (3.17):

(3.17)	y	bachgen	y	gwnaeth	y ci	ei	weld
	the boy		COMP	did-3SG	the dog	3MSG	see

'the boy that the dog saw' (Tallerman 1990:302)

[10]The ordering of these two strategies may depend on an assessment of the degree to which the two types of pronoun encode ϕ-features across languages.

[11]A Mann-Whitney U test gives us a significance level of $p < 0.005$, but this level may partially be due to the sampling technique.

(3.18) y papur roeddwn i'n edrych arno fo
the paper COMP-was-1SG I-PROG look at-3MSG it(3MSG)
'the paper that I was looking at' (Tallerman 1990:306)

3.6 Extending the explanation

The discussion in this chapter has led to the conclusion that a gradient hierarchy of processing complexity cannot on its own give rise to the cross-linguistic implicational hierarchy of accessibility to relativisation. Instead, a shifting competing motivations explanation is required. This inevitably gives rise to the question: can any implicational universal be explained without competing motivations? The rest of this chapter looks at this question for a few more cases, but any conclusions are rather speculative, opening up avenues for future research.

3.6.1 Simple extensions beyond syntax

The first two examples relate to fairly trivial processing/functional explanations in morphology and phonology. They should really be considered as simple illustrations of the way in which the method discussed in this chapter can be extended to non-syntactic domains.

Morphology It is well known (Greenberg 1963) that if a language marks gender distinctions in the first person, then it will mark gender distinctions in the second or third persons or both.

For gender marking: $1 \rightarrow (2 \vee 3)$

If the competing motivations approach is as general as the previous sections have suggested then we can make a direct analogy with the explanation for $O \rightarrow S$ and expect the following sorts of complexity differences:

1. First person gender marking is more complex than second or third person gender marking.

2. The lack of gender marking is in general more complex than gender marking.

On the other hand, the implicational universal above, has its *contrapositive* equivalent (see chapter 1):

For no gender marking: $(2\&3) \rightarrow 1$

This means that the other possible complexity differences should be:

1. The lack of second and third person gender marking is more complex than the lack of gender marking on first person.

2. Gender marking in general is more complex than the lack of gender marking.

Only these complexity pressures would give us something like the type graph in (3.8) and hence the implicational universal.

It seems that the latter possibility is the most likely one, especially given that marking gender by definition involves an increase in m-complexity over no gender marking, whatever persons are marked. How can we interpret the competing motivation in this case? Intuitively the hearer must, during parsing, map nominal expressions onto possible referents. The difficulty of this task in part relates to the amount of information about the referent that is encoded in the expression, so gender marking is useful inasmuch as it aids the mapping of signifier onto signified. It is likely, however, that gender marking is less important for first person expressions since the referent, at least for spoken language, is unambiguously given by context. Of course, one might wonder that this is not also true for second person expressions. This is only the case where there is only one possible addressee, however.

It seems then, that p-complexity increases when gender is left unmarked, especially on second and third person expressions, but conversely m-complexity increases when gender is marked on any expression. Again, the relative 'strengths' of these two pressures will vary dependent on the structure of the rest of the language as well as with context. So, for example, the difficulty of relating a referent to an expression depends not only on gender marking, but also the other types of morphological marking made available by the language. This is a direct analogue of the main case described in this chapter, and hence the implicational universal is expected.

Phonology The second example in this section relates to the diachronic tendency for ends of words to 'erode' over time. The argument is taken from Berg (1995), although it has been recast somewhat to highlight the similarities between it and those in this chapter.

Berg argues that there is an asymmetry in the perception of beginnings and endings of words. Hearers receive words as a sequence of acoustic events that run from some point in time t_1, the beginning of the word, to a later point t_2, the end of the word. At what point do hearers *recognise* the word? It is possible that the word will not be recognised until t_2, in other words after the last segment of the word. However, Luce (1986) (cited in Berg 1995) has shown that most English words over 5 segments long are *unique* before this point. This means that, even if a hearer was listening to words in isolation, then he could recognise the majority of them at time t_r after t_1 and before t_2. Since words are more likely to occur within an environment that facilitates their recognition, it is more likely that t_r will be even earlier, coming before the word's 'uniqueness point'.

Any perturbations in the prototypical sound of a word will clearly adversely affect its recognition. Given that a word is likely to be recognised before its end, however, any such distortions after t_r in the word will be of little consequence to the hearer. On average, then, there is a processing cost associated with distortions (especially reductions) in the phonological structure of a word, and this cost declines along the length of the word. The actual costs will differ at each occasion of use, since the recognition point is dependent on context. On the other hand, there is a natural tendency for phonological reduction regardless of the position in the word associated with articulatory effort.

These two pressures on the phonological structure of words lead to the asymmetry in diachronic erosion. The explanans in this case is somewhat different from others we have looked at in that it cannot be stated as an implicational universal (although "if a word has been eroded at a point before its end, then it will have been eroded at its end" gets close). This is because we are dealing not in discrete types and positions on a hierarchy, but rather with a continuum both in terms of extent of reduction, and average position on words. The shifting competing motivations approach, however seems to apply well in this example.

3.6.2 Word order revisited

One of the major problems relating to a generalised competing motivation approach is how it can be combined with the explanation for word order universals in the last chapter. In other words, does an explanation based on EIC admit the possibility of other motivations in conflict?

Matrix disambiguation

One of the implicational universals covered by Hawkins (1990, 1994a) seems to pose a problem for the EIC approach:

$$VO \rightarrow CompS$$

This means that almost all VO languages are Comp-initial in S′, whereas OV languages are found that are both Comp-initial and Comp-final. Early Immediate Constituents leads us to expect the MNCCs of the ICs of the verb phrase to be arranged close together, minimising the size of the constituent recognition domain. Here the MNCCs of the VP are V and Comp, so the expected optimal orderings are: $_{VP}[V\ _{S'}[Comp\ S]]$ and $_{VP}[_{S'}[S\ Comp]\ V]$. This is also what we would expect from Dryer's (1992) branching direction theory. Both V and Comp are non-branching categories, so in the unmarked case should order on the same side as their branching counterparts.

What about the other order predicted by the universal: $_{VP}[_{S'}[Comp\ S]\ V]$? This is not a problem for Dryer, since the BDT has nothing to say about *implicational* universals such as these, only about the (parametric) correlations between non-branching categories and verbs, and branching categories and objects — a correlation that is born out in this case since CompS is significantly more common amongst VO languages than OV, and SComp is *only* found in OV languages:

> "... there seems to be little question that this is a correlation pair. While both initial and final complementisers are found in OV languages (cf.

> Dryer 1980, Hawkins 1990, 225), complementisers in VO languages seem invariably to be initial; in fact, it may be an exceptionless universal that final complementisers are found only in OV languages. If so, then final complementisers are clearly more common in OV languages than they are in VO languages, and complementisers are therefore verb patterners, while the Ss they combine with are object patterners." (Dryer 1992:101–102)

In other words, the occurrence of OV&CompS is left unexplained.

Hawkins, on the other hand suggests two possible explanations for this asymmetry. The first (Hawkins 1990), based on the Minimal Attachment principle (e.g. Frazier 1985; Frazier & Rayner 1988) will not be discussed here. The second explanation (Hawkins 1994a:§5.6.1) is to do with the functions of a category like Comp other than mother-node construction. Consider the problems that the order $_{VP}[_{S'}[S\ Comp]\ V]$ might cause a hearer. Because the initial category in S′ is S, there is a potential for garden-pathing here; only once the complementiser is reached does the subordinate nature of the preceding clause become apparent (see, e.g. Clancy *et al.* 1986 for experimental evidence relating to similar examples involving relative clauses). There is a potential advantage, then, for "matrix disambiguation" immediately the S′ is encountered.

The following list sets out the parsing preferences of the various language types:

1. VO&CompS: Good for EIC, immediate matrix disambiguation.

2. VO&SComp: Bad for EIC, non-immediate matrix disambiguation.

3. OV&CompS: Bad for EIC, immediate matrix disambiguation.

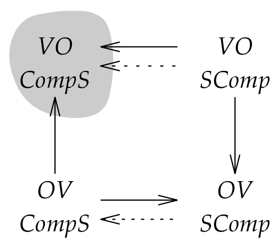
4. OV&SComp: Good for EIC, non-immediate matrix disambiguation.

All the occurring language types either have immediate matrix disambiguation or are good for EIC. Only the non-occurring type is both bad for Early Immediate Constituent recognition and does not immediately disambiguate between matrix and subordinate clauses.

This seems to be a neat explanation for the asymmetry. Indeed, it also seems to follow the structure of the explanation for the AH, in that the two pressures on parsing are in competition for OV languages.[12] One potential problem with it is that there is no definite reason why matrix disambiguation should be singled-out as such an important factor in parsing. Why do we not find a preference for immediate genitive disambiguation, for example? Although, the matrix/subordinate distinction is particularly significant in language (Hawkins 1994a:325), I believe this weakens the explanation somewhat.

A more crucial problem with this seeming competing motivations explanation can be understood with a type-graph:

(3.19)



On this graph, the solid arcs correspond to EIC motivated changes, and the dotted ones to change motivated by immediate matrix disambiguation. It is immediately obvious that this is not the same as the graph (3.8). The shaded area corresponds to the universal predicted by the type graph theory in this chapter: VO&CompS. This language type is a 'sink' since there are arcs leading into it, but none leaving it. Essentially, if this language type is the best possible for both EIC *and* immediate matrix disambiguation, then why shouldn't all languages end up being that type?

It is not clear what the ultimate solution to this problem might be. One might argue that there are other pressures, as yet unconsidered, in the word order domain that will mitigate the situation, particularly since EIC shows that the order of all constituents are related if they can appear in the same utterance (i.e. pressures on some other constituent's order may indirectly affect the type graph above). Alternatively, it may

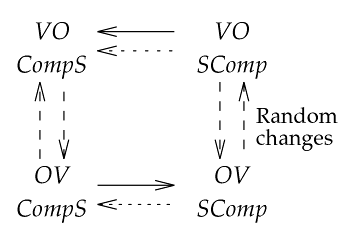
[12] The competition here is not between speaker and hearer, but rather 'within' the hearer. There is nothing in principle in the theory to rule this out, however.

have something to do with the origin of variation. This has so far been considered to be random with respect to the functions being examined (see McGill (1993) for discussion); some of the arcs in the type graph may be 'pruned' if this wasn't the case.

A different suggestion will be put forward here. So far it has been assumed that selection takes place over utterances. This means that, if an utterance proves hard to parse, then it does not form part of the trigger *and none of the information about word order that it contained will be presented to the LAD*. This seems a sensible stance to take in the absence of decisive experimental evidence about the contents of the trigger experience. On the other hand an alternative hypothesis might be more realistic. If a structure contains an embedded constituent that is hard to parse this does not necessarily mean that the branching direction of the superordinate structure cannot be adduced. In the structures being considered here, it may be possible to tell if the verb follows or precedes its object even if the order of Comp and S makes the recognition of the VP difficult. Furthermore, it is likely that there will be more examples of verb-object order in the rest of the utterances presented to the child that will not involve subordinate clauses. This means that EIC considerations might play their role in the selection of variant orders of Comp and S in S′, but not in the selection of variant orders of verb and object, or at least not to the same degree.

If this is the case we can redraw the type graph (3.19) to include changes between OV and VO which we can assume are random (i.e. not affected by the order of S and Comp):

(3.20)



This graph does not in effect rule out *any* language types, but the type VO&SComp is predicted to be less common (only one arc leads into it, but two lead out) and this becomes more marked if the languages retain their verb-object order for longer than the order of their complementiser and subordinate clause (i.e. if the changes in the former are rarer than the latter).

The same approach may also solve a problem with the EIC pointed out in Kirby (1994a:204–206). Wherever there are multiple MNCCs for a particular mother node, there will be a preference for languages that order MNCCs to the left. For example, given that Det and N are both MNCCs for NP, the first MNCC of an NP made up of Det and N in any order will always be the first word in that NP. In both the constructions $_{VP}[V\ _{NP}[Det\ N]]$ or $_{VP}[V\ _{NP}[N\ Det]]$, the constituent recognition domain will be the optimal two words. (Incidentally, this means that the order of determiner and noun should not be predictable from the order of verb and object. This is indeed the case (Dryer 1992).) For verb-final constructions $_{VP}[_{NP}[Det\ N]\ V]$ or $_{VP}[_{NP}[N\ Det]\ V]$ the CRD cannot be this short since it will always proceed from the first word of the NP to the verb.

This suggests that head initial languages will always contain constructions that are easier to parse than their head final counterparts, and that a type graph of all possible word orders would inevitably lead to a consistently head initial sink. If, however, selection does not take place at the level of the utterance, and the 'global' frequency of different constructions is taken into account as suggested above, then it is possible that these small differences will not have this effect.

Of course, this is only a tentative suggestion, the implications of which require testing against historical data and with further simulation work. One fruitful avenue of research would be to look at the influence of parsing on creolisation, where we might expect the availability of a huge range of input variation to allow for sampling from the complete range of possible orderings. Hence, the prediction would be that the set of word order types found in creoles is more like the set of ultimately optimal types for principles like EIC.

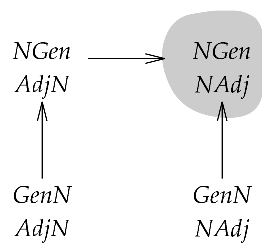
The prepositional noun-modifier hierarchy

The type-graph approach introduced in this chapter highlights some problems with the explanation for the prepositional noun-modifier hierarchy (repeated below) given in the last chapter. These problems are far from solved, but once again I will suggest some possible areas where a solution might be found.

> In prepositional languages, within the noun-phrase, if the noun precedes the adjective, then the noun precedes the genitive. Furthermore, if the noun precedes the genitive, then the noun precedes the relative clause.

For simplicity, let us consider only one of the implicational universals underlying the hierarchy: $GenN \rightarrow AdjN$ (for prepositional languages). The explanation given was that genitives were typically longer than adjectives, and in a structure $_{PP}[P\ _{NP}[Mod\ N]]$ the longer the modifier the worse the corresponding EIC metric. This means, if you like, that there is pressure for a language with prenominal genitives and prenominal adjectives to change its genitive-noun order first. This is backed up by the simulation results in figure 2.13. This means that the type graph for this universal is:

(3.21)



Once again, the problem is clear: the optimal type is a sink, so why do the other types occur? The same thing can be said about the universals $RelN \rightarrow GenN$ (figure 2.14) and $RelN \rightarrow AdjN$. This is the same problem that was faced trying to explain the accessibility hierarchy. In that case the problem was solved by invoking competing motivations, with shifting background conditions. But in the present case, it is hard to see what competing motivation there could be.

The danger with (3.21) is that it overly simplifies the situation. The mirror image universal applies for postpositional languages: $NGen \rightarrow NAdj$ (for postpositional languages).[13] In other words, if the adposition order of a language changes then there

[13] Interestingly, there is not an equivalent universal $NRel \rightarrow NGen$ for postpositional languages. This can be explained in terms of matrix disambiguation, as in the last section. In other words, there is a

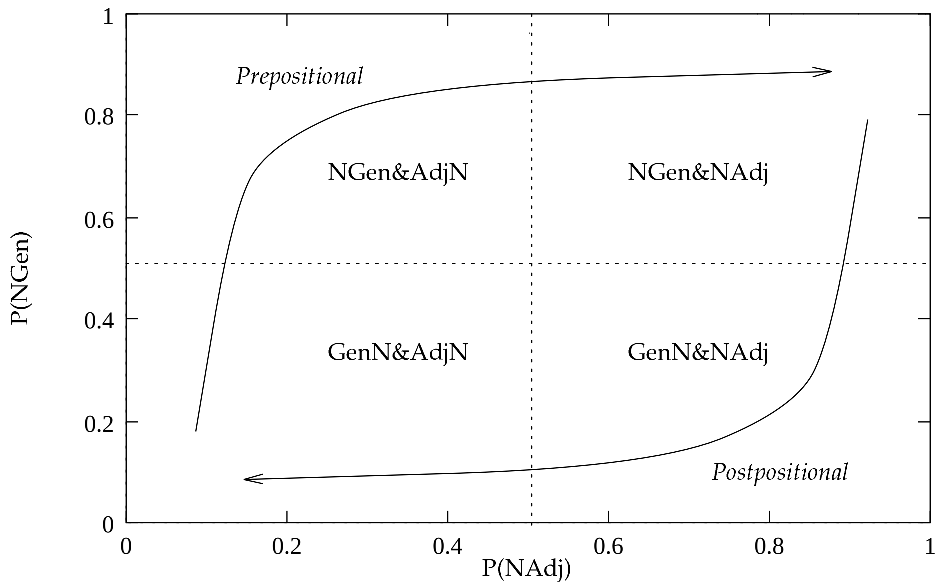


Figure 3.6. The predicted flow of languages through NAdj/NGen space.

will be a markedness reversal (see §2.5). This is because the preferred position, for the EIC, of the modifier is prenominal in postpositional structures. If a speech community is in the sink in (3.21) then such a markedness reversal will tend to start the language moving again (see figure 3.6).

There are some problems with this suggestion which need further research. For example, if the adpositional order were to change during the transition between the 'harmonic' types AdjN&GenN and NAdj&NGen, then a non-predicted type would arise (i.e. AdjN&NGen&Postp, or NAdj&GenN&Prep). If the rate of adpositional order changes is low enough, then this would not arise; this needs to be tested against historical data. This leaves us with the counter-intuitive position that adposition order only changes when it is maximally inefficient for it to do so (e.g. when a prepositional language changes to a postpositional one with consistent NMod order). Again, however, the selection of adpositional order may well be independent of

left/right asymmetry in the order of noun and relative clause with noun-initial being preferred in order to avoid garden pathing (Clancy *et al.* 1986).

modifier order for the reasons given in the previous section.

These problems aside, I hope that this brief discussion has highlighted the importance of looking carefully at the mechanism linking functional pressures with cross-linguistic universals, before making the assumption that they can be directly correlated.

3.6.3 The agreement hierarchy

The agreement hierarchy of Corbett (1983) is another example of a universal that we might attempt to explain using the principles set out in this chapter. The hierarchy predicts the distribution within and across languages of *syntactic* and *semantic* agreement between a controller and a target. These examples should make this terminology clearer:

(3.22) a. This team played cricket.

b. *These team played cricket.

(3.23) a. The team plays cricket.

b. The team play cricket.

(3.24) a. ?The team won the game it played.

b. The team won the game they played.

(3.22a) and (3.22b) show that *team* is syntactically singular. *Team* is the controller here, and the attributive modifier *this* agrees with it syntactically. In (3.23a) the predicate *plays* also agrees with the controller syntactically. (3.23b) is an example of another possibility: 'semantic' plural agreement. This option is also available for the anaphoric pronoun in (3.24b). In fact, some speakers find (3.24b) better than (3.24a) where there is syntactic agreement between the controller and the anaphoric pronoun target.

Corbett (1983) looks at syntactic and semantic agreement in Slavic languages in some detail, and proposes the Agreement Hierarchy:

attributive modifier>predicate>relative pronoun>personal pronoun

> "For any controller that permits alternative agreement forms, as we move rightwards along the Agreement Hierarchy, the likelihood of semantic agreement will increase monotonically. In absolute terms, if semantic agreement is possible in a given position in the hierarchy it will also be possible in all the positions to the right. In relative terms, if alternative agreement forms are available in two positions, the likelihood of semantic agreement will be as great or greater in the position to the right than in that to the left." (Corbett 1983:10–11)

The English examples above correspond to three positions on the hierarchy: attributive modifier, predicate and personal pronoun. Many cases in Slavic languages are given by Corbett as evidence for the hierarchy including ones in which the relative pronoun agrees with its controller.

For example, the Czech noun *děvče* "girl" is syntactically neuter singular. Semantic feminine agreement is possible with personal pronouns (3.27): (data from Vanek 1977 cited in Corbett 1983, 11-12)

(3.25) to děvče se vdalo
that(neut) girl got married(neut)

(3.26) najmula jsem děvče, které přišlo včera
hired did girl which(neut) came yesterday
"I hired the girl who came yesterday."

(3.27) to děvče přišlo včera, ale já jsem je / ji nenajmula
that girl came yesterday but I did it(neut) / her(fem) not hire

Another example involves the Russian noun *vrač*, which is syntactically masculine, but can enter into semantically feminine agreement relations when referring to a woman. This is true generally for Russian nouns which refer to people belonging to certain professions (data from Panov 1968 cited in Corbett 1983, 31–32):

(3.28) a. Ivanova, xorošij vrač
Ivanova (is) a good(masc) doctor

b. Ivanova, xorašaja vrač
Ivanova (is) a good(fem) doctor

(3.29) a. vrač prišel
the doctor came(masc)

b. vrač prišla
the doctor came(fem)

The percentage of informants selecting feminine agreement in a questionnaire study was higher for the predicate targets (3.29a–3.29b) than for the attributive targets (3.28a–3.28b).

I propose that an explanation for this hierarchy can take exactly the same form as that proposed for the accessibility hierarchy. Firstly, we need a definition for the syntactic complexity of the agreement relations in the above examples:

Structural complexity of agreement The structural complexity is calculated by counting the nodes in the *agreement domain*.

Agreement domain The agreement domain consists of that subset of nodes dominated by the lowest node dominating both target and controller that structurally integrate the target.

Structural integration of a node X in C The set of nodes which structurally integrate X in C are:

- all nodes dominating X within C (including C itself)
- all or some sisters of X (depending on surface coding conventions)
- all sisters of the nodes dominating X within C

This is an exact parallel of the definition of relativisation domains except that the node C differs depending on the target (i.e. for attributive modifiers C will be D′, for relative pronouns DP, and for predicates and many personal pronouns the node C will be IP). This is to be expected since structural complexity is a general measure of

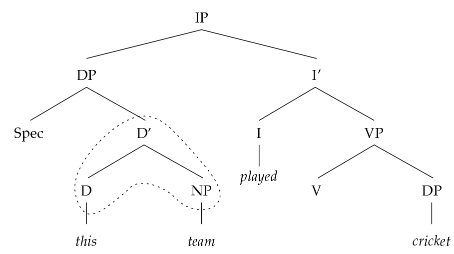


Figure 3.7. Attributive agreement.

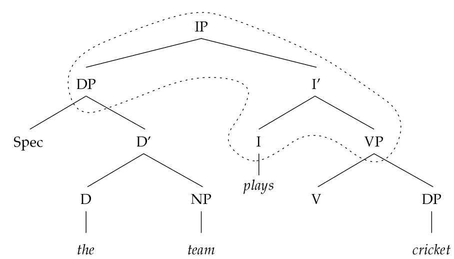


Figure 3.8. Predicate agreement.

tree-complexity relative to some psycholinguistic operation. Here, the assumption is that *syntactic* agreement involves some unique psycholinguistic operation.

The tree structures for (3.22–3.24) are shown in figures 3.7–3.9, with the agreement domains circled. This shows a clear increase in the structural complexity of agreement for the different targets. More generally, the structural templates in the following list show that the positions of the agreement hierarchy correspond to a hierarchy of structural complexity (where an *a* subscript indicates agreement):

attributive	$_{DP}\{N_a \ldots Mod_a\}$
predicate	$_{IP}\{_{DP}\{N_a \ldots\}\ _{Pred}\{V/Adj_a \ldots\}\}$
relative pronoun	$_{DP}\{N_a \ldots _{CP}\{wh_i\ _{IP}\{t_{i/a} \ldots\}\}\}$
anaphoric pronoun	$\{N_a \ldots Pron_a\}$

These structural templates are intended to show that the range of possible structures that could be involved for each target involves an increasing syntactic ‘distance’ in

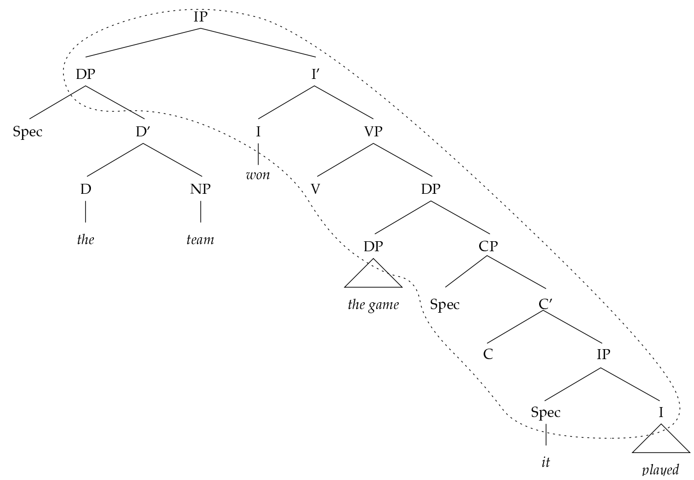


Figure 3.9. Personal pronoun agreement.

terms of agreement domains. The domain for the attributive modifier will typically involve only the sisters of N. The domain for the predicate will include all the nodes dominating the predicate within S, their sisters and the nodes dominating N in NP. The domain for relative pronoun only fits into the hierarchy in this place on the assumption that its trace carries agreement features in some way, extending the agreement domain arbitrarily deep within the clause. In this view, the target is not the relative pronoun itself, but the whole chain including the wh-element and the co-indexed trace. The potential domain for anaphoric pronouns is the largest since the target and controller can be in different matrix clauses.

Now that a tentative definition of the structural complexity of agreement has been defined we are left with exactly the same problems as with the accessibility hierarchy earlier in this chapter. It is not good enough to simply define a structural complexity hierarchy and assume it directly gives rise to a cross-linguistic hierarchy because one needs to explain why not all languages opt for minimum complexity, i.e. the top end of the hierarchy. The competing motivation in this case is probably something to do with the role of agreement in parsing (Hawkins 1994a:366–373). Essentially, syntactic ϕ-features can act as extra (redundant) information about the structure of the parse.

In Hawkins' terms, they can increase the construction potential of a node in parsing. Exactly how this interacts with the principles underlying EIC needs some working out, but the basic point is clear: syntactic agreement gives the hearer information at one point in the parse about other nodes in the parse.

In summary, syntactic agreement has a cost associated with structural complexity of agreement domains, and this complexity increases along the agreement hierarchy. In conflict with this is a parsing preference for redundancy of information that is provided by syntactic agreement. Just as similar competing motivations cause the accessibility hierarchy to emerge, the agreement hierarchy should emerge with these pressures in place. Clearly, there are many details of this putative explanation that need to be worked through. For example, does the definition of agreement domains make the correct performance predictions? Is the preference for small agreement domains a speaker- or hearer-driven pressure? These questions and others that arise from the specific approach to implicational universals expounded in this chapter will have to wait for future research.

Instead, the next chapter turns to cases where the general functional approach appears to fail and asks the question: what are the limits of adaptation?

Chapter 4

The limits of functional adaptation

In the previous two chapters I have argued that various universals, both parametric and hierarchical can be explained by examining the way in which processing complexity affects the transmission of language through the Arena of Use.[1] Computer simulations of language as a complex adaptive system have been useful in demonstrating the validity of this approach, as well as highlighting the limitations of previous explanations for hierarchies. The overall goal has been to solve the problem of linkage by enriching the structure of the Arena of Use proposed by Hurford (1990).

Now that a workable solution for the problem of linkage has been put forward, and given the stipulation that shifting competing motivations are required to explain hierarchical universals, it might be tempting to return to the situation outlined in the first chapter and accept any explanation that equates processing complexity and cross-linguistic distribution. Specifically, can we not now expect a cross-linguistic asymmetry whenever there is a psycholinguistic asymmetry?

This chapter looks at this question and answers it in the negative. It discusses some examples where a processing asymmetry does not give rise to a cross-linguistic asymmetry, and others where linguistic asymmetries appear to be related to the 'wrong' processing asymmetries. These results, then, appear to be fatal to the selection approach and, arguably, functional approaches in general. However, understanding

[1] A short version of the first half of this paper appears as Kirby 1994b.

these anomalies properly involves a radical reassessment of the role of innateness in explanation, and offers an interesting challenge to those trying to uncover the nature of universal grammar.

4.1 Another selection pressure on relative clauses

In the discussion on the accessibility of noun phrases to relativisation in the previous chapter, relative clauses where categorised according to the grammatical function of the trace, or resumptive pronoun, within the subordinate clause. So, for example, the following sentences exemplify the first two positions on the hierarchy:

Subject: *The man who found me saw Ruth*

Object: *The man who I found saw Ruth*

Any such categorisation is based on choices about what is relevant to typology, and what is not. It could be argued that a categorisation on the basis of the number of phonemes in the subordinate clause is equally valid, for example. It is unlikely that this would illuminate any particularly interesting cross-linguistic facts, however. In this section, the categorisation of relative clauses will be enriched by taking into account the grammatical function of the head noun in the matrix clause. This is also an available option and, as will be seen, it is commonly discussed in the psycholinguistic literature.

If our attention is restricted solely to the grammatical functions subject and object the following four categories of relative clause are distinguished:

Matrix subject, subject relative: *The man who found me saw Ruth*

Matrix subject, object relative: *The man who I found saw Ruth*

Matrix object, subject relative: *Ruth saw the man who found me*

Matrix object, object relative: *Ruth saw the man who I found*

A notation of the form X^Y will be used to signify a relative clause whose head noun has the function X in the matrix clause and whose trace, or resumptive pronoun, has the function Y in the subordinate clause.[2] The four sentences above are examples of S^S, S^O, O^S and O^O respectively.

One selection pressure on these relative clause types has been reviewed already. A study by Keenan & Hawkins (1987) looks at native English speakers' 'mastery' of relative clauses dependent on the function of the trace in the subordinate clause, using a repetition task. In their work Keenan and Hawkins make no mention of matrix function so we can characterise their results as follows on the assumption that their results should be generalisable to all relative clauses:

Accessibility $\{S^S, O^S\} > \{S^O, O^O\}$

The first experiments on the role of matrix function *and* subordinate function were carried out by Sheldon (1974). She used an enactment task with English-speaking children and showed that relative clauses were easier to process if the matrix function of the head matched the function of the trace in the subordinate clause. The results of this study, then, are:

Parallel function $\{S^S, O^O\} > \{O^S, S^O\}$

This result has proven hard to replicate (MacWhinney & Pleh 1988) and many studies have been carried out that give other rankings of structures in English. For example, DeVilliers *et al.* (1979) gives the results $\{S^S, O^S\} > O^O > S^O$ with a similar enactment task. Clancy *et al.* (1986:252) summarise the results of Sheldon (1974) and Tavakolian (1981) for their five-year-old subjects as giving evidence for $S^S > O^O > O^S > S^O$, which is in accord with their own study of Korean.

MacWhinney (1982); MacWhinney & Pleh (1988) review nine different enactment studies and note that "the results show remarkable consistency for the pattern $S^S > \{O^S, O^O\} > S^O$" (MacWhinney & Pleh 1988:117). They also cite studies of French and German (Kail 1975; Sheldon 1977; Grimm *et al.* 1975) that lend support to this ranking.

[2]The notation used in the literature is simply XY. This is avoided since, in the previous chapter, language types were signified using a similar notation. So, for example, SO signified a language type allowing both subject and object relatives. S^O, on the other hand, means a relative clause such as *the man who I found saw Ruth*.

Their own study of Hungarian also bolsters this ranking, at least for unmarked word orders.

Clearly, this is a controversial area, and many different factors have been proposed to account for the rankings. However, the results given above, although appearing to be in conflict, are not inconsistent with an interaction of both parallel function and accessibility. To see this, consider the two possible combinations of these factors. Either accessibility will be a more important factor than parallel function or vice versa:

Accessibility>Parallel function $S^S > O^S > O^O > S^O$

Parallel function>Accessibility $S^S > O^O > O^S > S^O$

All the rankings discussed so far are compatible with one of these possibilities (in other words, there are no predicted differences in any of the results that are not also predicted by one of the two rankings above). It is quite possible that *both* of these rankings are correct, and other factors relating to particular experimental materials such as the sentences under investigation mean that either accessibility or parallel function becomes the more important factor. If this is the case then over all possible relative clauses the ranking would be:

Accessibility=Parallel function $S^S > \{O^S, O^O\} > S^O$

This is the same as the ranking of MacWhinney & Pleh (1988), although they do not argue for a combined accessibility/parallel function account of their results.

Before continuing, it should be pointed out that there is a methodological difference here. Accessibility has been given support by Hawkins' independent complexity theory as discussed in the previous chapter, whereas parallel function (or any other possible determinant of processing difficulty) is not supported in this way. This might suggest that accessibility is after all the only factor influencing relative clause complexity. The problem with this is that it fails on its own to predict (although it is consistent with) the psycholinguistic results, particularly the result on which there is least disagreement: that S^O relatives are harder to process than any others. It is not easy to work out what other universal principles are in operation, but clearly there is something more than accessibility at work. Let us assume for the moment that parallel function acts as a selection pressure in the arena of use.

4.2 A failure of the functional approach

In the previous chapter a competing motivations explanation for the accessibility hierarchy was put forward that related the processing asymmetry $S > O$ with the cross-linguistic asymmetry $O \rightarrow S$, given a competing dispreference for non-relativised alternatives. In the notation given above this means that $\{S^S, O^S\} > \{S^O, O^O\}$ gives rise to $(O^O \vee S^O) \rightarrow (S^S \vee O^S)$. In order to test any such predicted universal, we can re-write the implication as $(S^S \vee O^S) \& \neg (O^O \vee S^O)$. The language types that we expect to find if accessibility influences the selection of relative clauses are therefore:

1. $S^S \& \neg O^O$
2. $S^S \& \neg S^O$
3. $O^S \& \neg O^O$
4. $O^S \& \neg S^O$

As discussed in the previous chapter, Keenan & Comrie's (1977) accessibility hierarchy explicitly states that all these language types exist:

> "For each position on the AH, there are possible languages which can relativise that position with a primary strategy, but cannot relativise any lower positions with that strategy." (Comrie & Keenan 1979:653)

In principle there is no reason why any other asymmetrical pressure on the processing of relative clauses should not also give rise to an implicational universal. In other words, there is nothing in the logic of the competing motivations explanation that rules out parallel function as a further factor in determining the p-complexity of RCs. Following the same logic as above, the influence of parallel function $\{S^S, O^O\} > \{O^S, S^O\}$ should give rise to the universal $(O^S \vee S^O) \rightarrow (S^S \vee O^O)$. This can be re-written as a conjunction: $(S^S \vee O^O)\&\neg(O^S \vee S^O)$. Evidence for parallel function cross-linguistically should come as the following language types:

1. $S^S\&\neg O^S$
2. $S^S\&\neg S^O$
3. $O^O\&\neg O^S$
4. $O^O\&\neg S^O$

The second type corresponds to the second type giving evidence for accessibility and turns up as Iban, for example. The first, third and fourth types have not been found (although see the following section for apparent counter-evidence).

There is therefore no currently available evidence for parallel-function showing up cross-linguistically (although proving that some language type does not exist is impossible). Perhaps the problem is that the processing pressures are being considered in isolation, whereas we have argued that a combination of accessibility and parallel function is acting on the processing of relative clauses. The complexity hierarchy $S^S > \{O^S, O^O\} > S^O$ should give rise to the implicational universals:

$$S^O \rightarrow (O^S \vee O^O)$$
$$S^O \rightarrow S^S$$
$$(O^S \vee O^O) \rightarrow S^S$$

In turn these can be re-written as conjunctions:

$$(O^S \vee O^O)\&\neg S^O$$
$$S^S\&\neg S^O$$
$$S^S\&\neg(O^S \vee O^O)$$

The predicted types are therefore:

1. $O^S\&\neg S^O$
2. $O^O\&\neg S^O$

3. $S^S \& \neg S^O$

4. $S^S \& \neg O^S$

5. $S^S \& \neg O^O$

Once again, some of these types do occur (1, 3 and 5), but these are simply the ones that we have evidence for from the work on the accessibility hierarchy. The critical types regarding the added influence of parallel function are 2 and 4, and there is currently no evidence for the existence of these language types.

This poses serious problems for the functional approach put forward in this thesis so far. There is nothing in the theory that can explain why accessibility has cross-linguistic implications, but parallel function has not. It seems that the explanations put forward here suffer from being ad hoc, a common criticism of functional explanations (see, e.g. Lass 1980).

4.3 Innate constraints on adaptation

The failure of parallel function to show up cross-linguistically seems to be a fatal blow for functional explanations but this is because we have so far only been looking at one side of the coin as regards the adaptive nature of language. The map of transformations in the cycle of acquisition and use from chapter 2 is shown again in figure 4.1. So far, we have only been concerned with the transformations **T1** and **T3** (production and parsing), treating the relationship between trigger and competence (**T4**) as a simple mapping. Recall that the simulations in the previous two chapters treated competence as a list of utterance types — individual grammars were 'acquired' by compiling such a list directly from the trigger experience. The only assumption that was made was that acquisition is an all-or-nothing process. In other words, the acquired competence does not directly reflect subtle frequency effects in the trigger (although marked variants can be acquired as marked variants having something like "foreign language status"). This is clearly a gross simplification of what is actually going on in acquisition, but it is justified inasmuch as we believe that the function

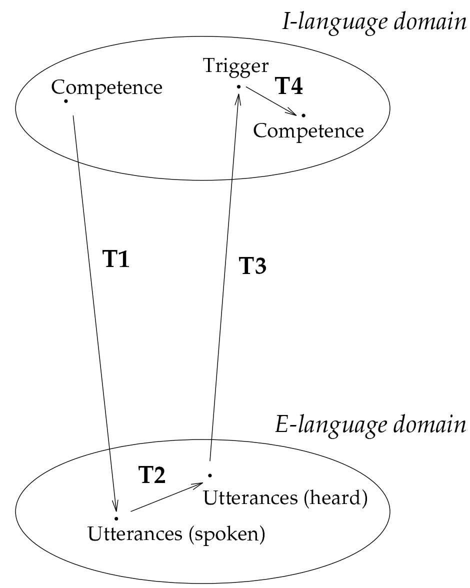


Figure 4.1. Transformations within and between *I*- and E-domains.

mapping trigger onto competence does not affect the viability of variants over time. Furthermore, though less obviously, it also rests on an assumption that the medium of representation of competence does not also affect variant viability.

It is quite possible that something about the process of acquisition distorts the distribution of variants in the trigger in more profound ways than assumed so far. This might be due to constraints imposed by the acquisition device, or it might be due to constraints imposed by the nature of competence itself. In other words, the structure of a grammatical meta-language may not in fact be able to accurately represent features of the trigger experience. If this were true then certain *constraints on adaptation* should be expected.

4.3.1 Constraints on adaptation in biology

Before going on to explore the implications of constraints imposed by acquisition or competence, it might be useful to look at a similar problem that crops up in another field of complex adaptive systems.

As was discussed in chapter 1, the adaptive nature of forms in the biological world has much in common with the adaptive nature of language. Both exhibit, to some extent, a striking 'fit' of form to function which inevitably leads us to look for an explanation of that form in terms of function. Although there are a number of crucial differences, the theory that links function and form in language proposed here has much in common with neo-Darwinian selection theory. Indeed, both areas have their generalised form in a theory of complex adaptive systems (Gell-Mann 1992; §2.2.3). It will be instructive, therefore, to look at a couple of cases of mismatches between form and function in biological evolution discussed by Gould (1983:147–165).

The non-occurrence of a form

Imagine you are an engineer attempting to design some mechanism for moving a machine efficiently over a flat surface. A good design would maximise the distance to work ratio of the machine. Given enough time it is likely that you would plump for a design that has been used by engineers time and time again to solve this very problem: the wheel.

Wheels are functional because they minimise friction when a body is moving over ground, and they stay with the body as it moves (unlike rollers). Although they are not as versatile as legs, for example, in terms of the terrain they can cross, the bicycle is a good example of the combination of the two that is amazingly effective at increasing the mobility and speed of a human being. Given that wheels are so functional — they are perfect examples of 'fit' between form and function — it is surprising that they are vanishingly rare in the biological world. Human beings are the only organisms with wheels, and even for us they are not part of our biological phenotype, but our "extended phenotype" in Dawkins's (1982) terms. In other words, we do not grow wheels, but have to fashion them from raw materials in our surroundings. Here then is an apparent failure of the theory of natural selection. The forms that occur across the biological kingdom do not live up to expectations; there is a mismatch between form and function.

The solution to this problem lies in the nature of wheels:

> ". . . a true wheel must spin freely without physical fusion to the solid object it drives. If wheel and object are physically linked, then the wheel cannot turn freely for very long and must rotate back, lest connecting elements be ruptured by the accumulated stress." (Gould 1983:160)

The problem for biological organisms is that the parts that make up the organism must be physically connected in order for nutrients to flow between them. As Gould points out, some of our bones are disconnected, but require a surrounding envelope of tissues preventing their free, or wheel-like, rotation.[3] It is impossible, then for biological wheels (as opposed to wheels made of non-living matter) to exist in the physical world due to a constraint on permissible forms.

> "Wheels work well, but animals are debarred from building them by structural constraints inherited as an evolutionary legacy. Adaptation does not follow the blueprints of a perfect engineer. It must work with parts available." (Gould 1983:164)

The occurrence of a non-functional form

As well as the possibility of an expected form not turning up in biology, Gould gives an example of an unexpected form that cannot be understood without looking at constraints on adaptation. The particular example may initially seem irrelevant to a thesis on language universals, however as we shall see the similarities between this and the case of parallel function in relative clauses is striking.

The external genitalia of the female spotted hyena are remarkably similar to that of the male of species (so much so, that medieval bestiaries commonly assumed that the hyena was androgynous). This unusual similarity begs an explanation, although the selective advantage to the female of appearing to be male are rather hard to understand. One attempt at an explanation suggests that the female genitalia evolved

[3]It turns out that there is an exception to this rule. *Escherichia coli* has flagella that act like propellers. They are able to escape the constraint on physical connection only because of their small size. Nutrients and impulses are conveyed between the separate parts by diffusion.

for use in a meeting ceremony, where typically more conspicuous structures would have an advantage in "getting the owner recognised". However, Gould points out:

> "Speculation about adaptive significance is a favourite ... ploy among evolutionary biologists. But the question 'What is it for?' often diverts attention from the more mundane but often more enlightening issue, 'How is it built?' " (Gould 1983:152)

Gould's argument runs that male and female hyena genitalia are similar because the embryological development of the structures follows the same course. In the genetically coded program for ontogenetic growth there is nothing that forces the female and male structures to differentiate.[4] The point is that we do not have to explain the existence of the occurrence of the female form — it is forced on the hyena by constraints on the pathways of embryological development.

4.3.2 Formal constraints on relative clauses

The examples from biology show that the adaptation of forms to fit some function can be limited by physical constraints on morphogenesis. This can mean that an expected form does not show up, and, more unexpectedly, that non-functional forms can exist. This means, as Gould argues forcefully, that it is not possible to simply equate function with form. Mismatches are the expected outcome of the system into which adaptive changes must be born.

For the hyena, the external sexual characteristics of the female are forced upon her by physical constraints on embryological development; they are a side-effect, if you like, of the existence of similar structures in the male of the species. Can a similar argument be used to explain why it is not possible to get a parallel function relative clause without also getting the non-parallel function equivalent? If so, the absence of the expected cross-linguistic asymmetry should not cause us to reject the functional

[4] Of course, it is not impossible for other similar organisms to have this differentiation coded in the genome (such as other species of hyena), however this entails reducing levels of hormones in the female of the species. Gould suggests that the high levels of the hormones in the female spotted hyena are adaptive in some other way.

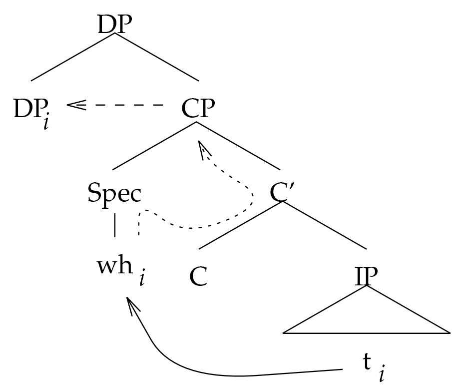


Figure 4.2. The structure of an abstract relative clause.

approach.

There must be something about the transformation from trigger experience to competence (the transformation mediated by the LAD) that forces the language user to acquire O^S relatives whenever S^S relatives are acquired, and S^O relatives whenever O^O relatives are acquired. The tree in figure 4.2 is the familiar formal representation of a relative clause. Although the details of this representation may vary slightly from one syntactic theory to another, the important characteristics for this argument are uncontroversial.

Firstly, notice that the trace dominated by IP, the wh-element in [Spec,CP] and the nominal head in DP are all related in some way. The interpretation of a relative clause such as *the man who I found* requires this. The relative pronoun *who* is related to the trace position (as can be seen by the *who/whom* distinction in certain registers of English); this is indicated by co-indexation. Furthermore, the head of the relative clause, *the man*, must be interpreted as being the logical object of the subordinate construction. The operator *who* in the relative clause is a referential expression standing in for the head noun, and sharing its ϕ-features. So, in many languages the relative pronoun agrees in person, number and gender with the head. This relation is also shown by co-indexation; in Principles and Parameters theory, the relationship between the head noun and the relative pronoun is actually assumed to be between the head noun and the 'chain' of wh-element and trace. Hence, all three are co-indexed.

The formal mechanisms by which these elements are related might vary from theory to theory. A standard assumption is that the wh-element has moved from the position of the trace in the subordinate clause. The head DP is in a "predication relation" with the CP, which inherits the trace of the wh-element in [Spec,CP] by some kind of generalised Spec-head agreement. Whatever the theory, there are two distinct operations going on: one relating trace and relative pronoun, and the other relating the head noun with the subordinate clause. It is unlikely that these two operations, predication and wh-movement, could be subsumed under one mechanism in any grammatical formalism.

Now, in general, there may be constraints on the operation of mechanisms such as predication and wh-movement. These may be universal in nature or language-specific, forming part of the native speaker competence for the language. If parallel function were to be realised cross-linguistically the language types $O^O\&\neg S^O$ or $S^S\&\neg O^S$ should show up. If such a language were to exist, it would fall to language-specific constraints on the operation of predication and wh-movement to express the grammaticality of the parallel function relatives and the ungrammaticality of the non-parallel function variants. However, in order to express exactly these grammaticality facts any constraint on predication would need to be dependent on information about wh-movement, or vice versa.

However, it is generally assumed that an operation like predication cannot be sensitive to the internal structure of the CP, and similarly wh-movement cannot be restricted on the basis of structure outside of the CP. These two operations in this structure are informationally encapsulated from one another. This means that, if these grammatical facts are mirrored in the LAD, the predicted language types are actually impossible to acquire or represent in the I-domain of figure 4.1. If a child acquires competence in response to a parallel function relative, then she cannot help but also acquire competence for the non-parallel equivalent. If the non-parallel function form is made ungrammatical, then the parallel function variant goes too.

The transformation **T3** will tend to filter out the forms that are more complex to

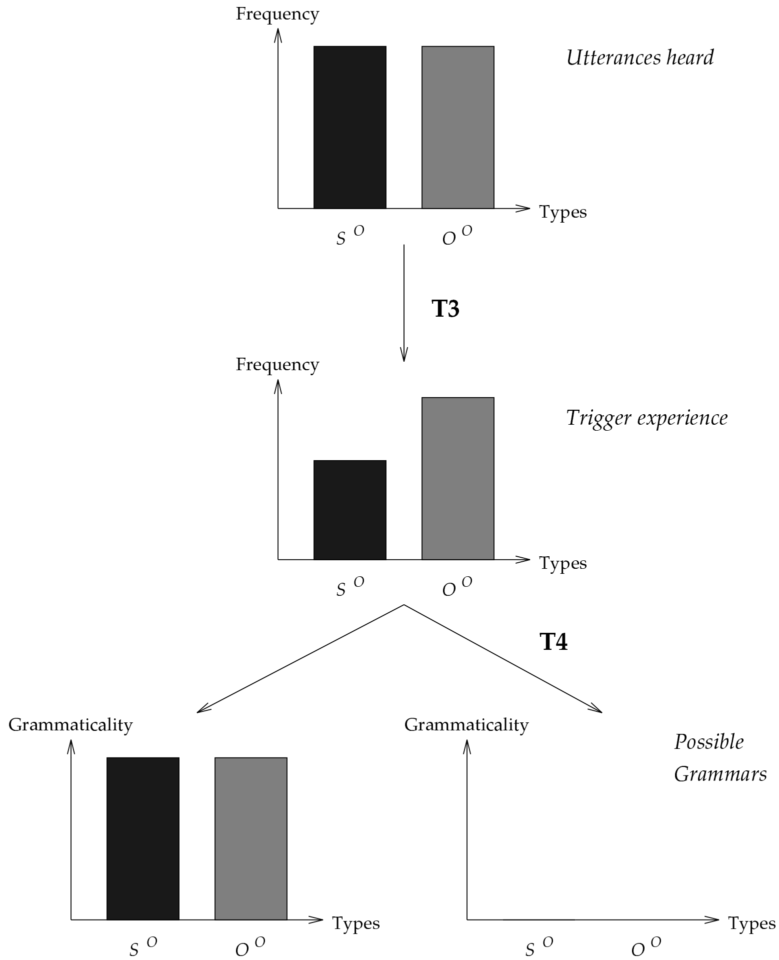


Figure 4.3. The possible pathways of S^O and O^O variants.

process. So, the theory of linguistic selection predicts that the proportion of, say, S^O variants relative to O^O variants that form part of the trigger should be lower than the proportion in the language data. However, given this differential distribution, the LAD (transformation **T4**) can only do one of two things: both variants can be made ungrammatical, or both variants can be made grammatical (figure 4.3). Even if *no* S^O variants made it into the trigger, they could still be acquired by the child. We might say that the S^O form is a *latent variant* in that it can be retained from generation to generation in the I-domain conceivably without ever being expressed in the E-domain.

Berg (1995) arguing from a rather different perspective also discusses the existence of latent variants. He assumes a representation of linguistic knowledge in the form of a highly connected network of units that each represent a specific linguistic feature. Whenever a unit in this “localised connectionist network” is activated, the activation spreads down links that may inhibit or excite the activation of neighbouring units. The activation of some set of phonological features in the production of a word, for example, might also excite neighbouring features without quite pushing them over an activation threshold. Berg argues that this kind of network can explain a large range of speech error data where neighbouring units are inadvertently activated. Even if speech errors do not occur, however, the variation exists “just below the surface” because of the structure of the network. This variation can be passed from generation to generation without ever showing up in the surface, but some perturbation in the language might bring the variant forms unexpectedly to the fore. This is another case where the structure of the acquisition device and representational medium of language means that the free selection of variants in the cycle of language acquisition and use is not always possible.

4.3.3 Some apparent counter-evidence

The argument put forward in the previous section seems to explain why the functional explanation for the accessibility hierarchy does not generalise to other processing asymmetries in relative clause constructions. The whole approach is put into jeopardy, however, if there are any counter-examples to the encapsulation of principles outlined above. This section introduces two cases where a language appears to have responded at least partially to pressures from parallel function.

Hopi relative clauses

Hale *et al.* (1977) note that “it would appear that Hopi exhibits a curious limitation on the accessibility of noun phrases to relativisation”. In matrix subject position, only subject relatives are acceptable; S^O relatives are ungrammatical (Hale *et al.* 1977:400–401)[5]:

(4.1) a. míʼ tiyóʼya ʼacáta-qa pákmīmīya
that boy lied-QA$_1$ cry
'The boy who lied is crying'

b. ʼitána mí-t tiyóʼya-t nˉíʼ tˉíwa:-qa-t hoóna
our-father that-OBL boy-OBL I saw-QA$_3$ sent home
'Our father sent home the boy whom I saw'

c. nˉí mí-t tiyóʼya-t ʼacáta-qa-t hoóna
I that-OBL boy-OBL lied-QA$_3$ sent home
'I sent home the boy that lied'

These examples are cases of an S^S relative (4.1a), an O^O relative (4.1b) and an O^S relative (4.1c) respectively. The "missing" relative clause type is shown below (Hale *et al.* 1977:402):

(4.2) * míʼ tiyóʼya nˉíʼ tˉíwa:-qa-t pay níma
that boy I saw-QA$_3$ already went home
'The boy whom I saw has gone home'

This is what would be expected if Hopi was responding to parallel function and accessibility. The complexity hierarchy that was argued for in section 4.2, $S^S > \{O^O, O^S\} > S^O$, should give rise to the universals $S^O \rightarrow (O^S \& O^O)$, $S^O \rightarrow S^S$ and $(O^O \vee O^S) \rightarrow S^S$, all of which are true for Hopi. Critically, the ungrammatical type appears to show that there is some mechanism whereby the position of the RC in the matrix can constrain the position that can be relativised. This is precisely what was claimed to be impossible in the previous section. It is important, therefore, that the properties of the Hopi relative clause are examined carefully.

The element *-qa* in the Hopi relative clauses seems to act as a relativisation marker that phonologically binds to the subordinate verb. In fact for other reasons Hale *et al.*

[5]The examples are taken directly from the cited source, except that the names of the suffixes on QA have been changed to numbers for clarity. The optional resumptive pronouns have also been omitted for clarity.

(1977) argue that this element is not simply a relativisation marker or complementiser, but acts as the head noun of the relative clause. The details of this argument are unimportant here, however. The crucial feature of the QA element is that it is assigned case in a rather peculiar way. In order to predict the morphological marking on the QA element, it is necessary to know whether the subject of the subordinate clause is coreferential with the subject of the main clause as well as the grammatical function of the trace in the relative clause. The three possibilities are (considering the singular only):

1. */-qa/*: coreferential subjects and subject relativisation
2. */-qa-y/*: coreferential subjects and non-subject relativisation
3. */-qa-t/*: otherwise

Only the first and third markings are apparent in the examples so far. The second type is exemplified by the O^O relative (Hale *et al.* 1977:400):

(4.3) n¯ı’ taávo-t n¯ı’ niína-qa-y sísk^{w}a
I rabbit-OBL I killed-QA$_2$ skinned
‘I skinned the rabbit that I killed’

This system of marking, although unusual, does not seem to help us explain the ungrammaticality of (4.2). Although S^S relatives are uniquely marked as *-qa*, there is nothing in the case marking system that reliably distinguishes the other three types.

Another feature of the morphological marking of the sentences above, is that all the non-subject noun phrases are marked with an oblique case ending */-t/*. Another possible oblique case ending is */-y/*, although this is not present in these examples. [6] The morphology of the second and third QA suffixes now looks very like */-qa/*+OBL, the choice of the two OBL forms being dependent on whether subjects are coreferential or not. In sentence (4.2), the noun phrase in subject position thus appears to terminate

[6]This is a considerable simplification of what is going on with the oblique in Hopi, although it does account for the data given here. See Hale *et al.* 1977, 394–402 for a more detailed account of Hopi relatives, based on traditional transformational assumptions.

with an oblique ending. However, this runs counter to the surface fact in Hopi that subjects are unmarked. The ungrammaticality of the S^O relative is therefore due to the incompatibility of the morphological rules that mark QA as oblique in S^O relatives and require subjects to be unmarked for case.

Further evidence for the "surfacy" nature of this constraint can be found by looking at the extraposed variant of (4.2) (Hale *et al.* 1977:402):

(4.4) mí' tiyó'ya pay níma, n¯í' t¯íwa:-qa-t
the boy already went home, I saw-QA_3
'The boy has gone home, whom I saw'

This variant on the S^O relative is grammatical in Hopi because the surface subject does not terminate with an oblique ending.

German free relatives

The second apparent counter-example comes from a sub-type of German relative clause constructions. The constructions in question are free, or headless, relatives — relative clauses lacking a head noun (see, e.g. Groos & van Riemsdijk 1979). Given that these constructions are rather different from the standard headed, restrictive relatives that we have been considering so far, it is not at all clear that the psycholinguistic results about relative processing complexity should apply. However, if these constructions exhibit a grammaticality constraint that involves the interaction of matrix function and subordinate function, then the argument put forward in the previous section about an innate limitation on the format of constraints will be put in doubt.

In fact, German free-relatives (at least for some native speakers) do exhibit just this kind of grammaticality pattern (Cann & Tait 1990:25):

(4.5) a. Ich muss wen du mir empfiehlst nehmen
I must who(acc) you to me recommend take
'I must take who you recommend to me'

b. * Ich muss wer einen guten Eindruck macht nehmen
I must who(nom) a good impression makes take
'I must take whoever makes a good impression'

c. * Ich muss wem du vertraust nehmen
I must who(dat) you trust take
'I must take whoever you trust'

The first sentence (4.5a) is an example of an O^O free relative, whereas (4.5b) is an example of an O^S relative, and is ungrammatical. There is not a simple constraint allowing O^O and not O^S, however, since (4.5c) is an O^O relative, but is also ungrammatical.

The pattern of grammaticality is predicted by comparing the morphological case of the relative pronoun, and the case assigned by the matrix verb. In (4.5a), the accusative relative pronoun matches the accusative case assigned by *nehmen*, but in the other examples there is a 'clash' between the case assigned by the verb and the morphological case of the relative pronoun. This does not explain what is going on in German, however, because the equivalent headed relatives are all grammatical:

(4.6) a. Ich muss den Mann den du mir empfiehlst nehmen
I must the man who(acc) you to me recommend take
'I must take the man who you recommend to me'

b. Ich muss den Mann der einen guten Eindruck macht nehmen
I must the man who(nom) a good impression makes take
'I must take the man who makes a good impression'

c. Ich muss den Mann dem du vertraust nehmen
I must the man who(dat) you trust take
'I must take the man who you trust'

The sentences (4.5a-c), then, seem to allow some way for information about the grammatical function of the trace to interact with information about the grammatical function of the complex noun phrase. This will be a problem for the theory if these free relatives are assigned a structure similar to that in figure 4.2.

Cann & Tait's (1990) analysis of these constructions suggests that this is not the case. The tree in 4.4 has the subordinate clause generated internal to the NP, rather than adjoined to DP. In this structure, the DP dominating the relative pronoun *wen* has moved from within the IP to [Spec,CP] as normal. This forms a chain (DP_i, t_i) which is assigned accusative case by *empfiehlst*. A further movement of *wen* to the head of

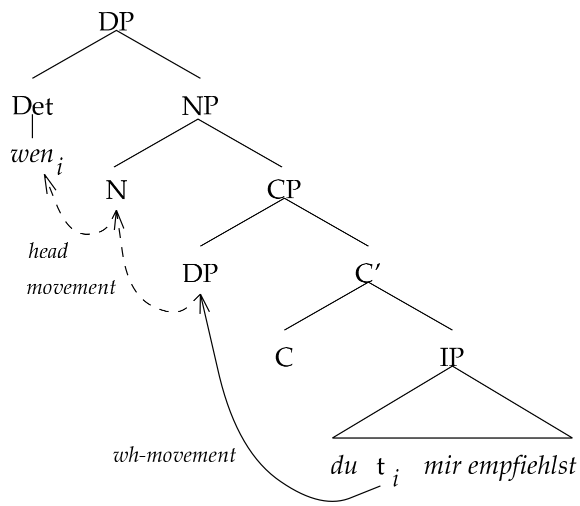


Figure 4.4. The structure of a German free relative.

the maximal DP is forced in the theory proposed by Cann and Tait. This movement is required to satisfy a phonetic-form licensing principle that has the effect of restricting the occurrence of phonetically-null nodes that do not form a part of a chain headed by a licensed node; in this case, the head of [DP,CP], the noun, and the head of the maximal DP.[7] Given this obligatory movement, the maximal DP inherits the case carried by its head *wen*. The category DP cannot be assigned contradictory feature values, so given that the two chains formed by the movement transmit the accusative case feature to the relative pronoun, the entire DP cannot be assigned anything other than accusative case by the matrix verb and yield a grammatical sentence.

For most speakers, the extraposed variants of (4.5) are grammatical (Cann & Tait 1990:25):

[7] This is not the place to discuss the details of Cann and Tait's phonetic-form licensing principle (PFLP), suffice to say that it is motivated by the need to constrain the set of functional projections that the language acquirer has to postulate by requiring every syntactic projection to have some phonological representation. It is interesting to note that this principle is very similar to Hawkins's (1994a) Axiom of MNCC Existence, which holds that every mother node must have a phonetically non-null constructor.

(4.7) a. Ich muss nehmen, wen du mir empfiehlst
I must take who(acc) you to me recommend
'I must take who you recommend to me'

b. Ich muss nehmen, wer einen guten Eindruck macht
I must take who(nom) a good impression makes
'I must take whoever makes a good impression'

c. Ich muss nehmen, wem du vertraust
I must take who(dat) you trust
'I must take whoever you trust'

Cann & Tait (1990) suggest that the structure of the relatives must be an adjunction structure $_{DP}[DP \;\; CP]$ (in other words, like the structure in figure 4.2). It cannot be the same structure as given for the free relatives in situ because moving the CP to the post-verbal position would leave the relative pronoun behind in the clause. Given the same structure as was put forward for non-free relatives, we expect the matching constraint to be impossible and hence the grammaticality of the sentences (4.7a-c).

The only question remaining is why Cann & Tait (1990) do not propose the adjunction structure for the non-extraposed free relatives (4.5), and instead opt for CP being generated internal to NP. The answer is rather technical, and only a flavour of it will be given here. Essentially, the phonetic form licensing principle requires that the empty DP_2 in the free relative construction $_{DP_1}[DP_2 \; CP]$ be *governed* by the relative clause, CP. Because DP_2 is part of an adjunction structure, the other segment of this structure DP_1 must also be governed. This is not possible if the CP is dominated by DP_1 as it is here. However, if the CP is extraposed then it is available as a governor of both segments of the DP. This problem of government is, on the other hand, not an issue with the structure given in figure 4.4 because the DP is not empty, and is therefore already phonetically licensed.

In summary, the German free-relative data, and the Hopi data appear to contradict the explanation given as to why parallel function does not show up cross-linguistically. A closer examination of the syntactic explanations for these language-specific phenomena reveals that this is not the case. The particular idiosyncrasies of the language and/or structure in question may allow the parsing preference to be realised grammatically after all — the status of these findings within the selection model will be considered later in this chapter. For the moment, the message should be that the architecture of grammar cannot be ignored in assessing the cross-linguistic effects of functional pressures. The next section further pushes this message home by uncovering a case where a weight distinction in processing crops up in a different form grammatically. In fact this is a case where the architecture of grammar (i.e. UG) means that the effects of selection are *mal*adaptive.

4.4 The English genitive

The prepositional noun-modifier hierarchy of Hawkins (1983) was discussed in chapter 2. One of the predictions made was that if a language has variable order at one position of the hierarchy, then it is likely that all modifiers higher on the hierarchy will order one way and the modifiers lower on the hierarchy will order the other way. Modern English exemplifies this nicely, with AdjN and NRel orders and variable order for genitives. The GenN genitive is the so-called Saxon genitive that has survived from Old English, formed by an inflectional suffix on the head noun. The 'Norman genitive' on the other hand is formed with a preposition *of* and appears only very rarely in Late Old English (Fischer 1992). The Modern language thus seems to be half way down the hierarchy:

$$Prep \rightarrow (AdjN > GenN > RelN)$$

The explanation put forward for this effect in chapter 2 relied on the idea that changes in the orders up the hierarchy happen in sequence and that as a change occurs variant orders may co-occur. In other words, we should expect the prenominal genitive in English to be on the way out, and the postnominal genitive to be on the increase as the language changes its modifier+noun orders in line with its adposition order. If we examine the order of genitives in Middle English, this prediction seems to be cast in doubt. In Middle English, by far the most common genitive construction was NGen, appearing about 85% of the time (Fischer 1992), with the prenominal genitive inherited from Old English as a minor variant. In Modern English, the prenominal genitive is clearly more than a minor variant. The situation is something like:

$$\text{(Middle English)}\ {}_{\text{GenN}}/\text{NGen} \rightsquigarrow \text{GenN}/\text{NGen}\ \text{(Modern English)}$$

The order that is dispreferred by EIC — the prenominal genitive — becomes *more* common in Modern English, so it looks less like the language is simply in transition between two points on the hierarchy. This is especially mystifying when Old English is considered since GenN was then the predominant order. The order of changes involves an introduction of preferred (from the parsing point of view) order in Modern English and the reduction in frequency of the dispreferred Old English order as expected. In Modern English, however, this trend is reversed with an increase in frequency of GenN. We are left with the question: why has the change turned around?

In order for the selection model to work, it was pointed out in chapter 2, the variants on which there is a processing pressure must be in competition. Kroch (1994) claims that the situation where grammatical variants are in competition is analogous to the situation where morphological doublets are competing for a paradigm slot. Where the two variants are functionally undifferentiated, then we expect language users to acquire one or other of the two variants at the expense of the other (although the other may exist as a form that is marked in some way, see §2.2.5). Notice, however, that the condition on competitive replacement by linguistic selection is that the variants are "functionally undifferentiated". Kroch (1994:15–16) gives Dutch adpositions as an example of a case where this condition is not met (examples due to Laura Joosten):

> "Dutch... has both prepositions and postpositions. In addition, a number of Dutch adpositions may be either prepositional or postpositional, with, however, a consistent difference in meaning. The prepositions are generally locative, while the postpositions are always directional. The examples below illustrate this behaviour [Kroch's (29)]:

(4.8) a. Ik fiets in de straat

I bike in the street
(locative only) 'My bike riding takes place in the street'

b. Ik fiets de straat in

I bike the street into
(directional only) 'My bike riding takes me into the street' "

Is it possible that the pre- and post-nominal genitives in Modern English have become functionally differentiated? This would explain why neither form is clearly a 'marked variant' having a kind of "foreign" status for native speakers as Kroch puts it, and it would also explain why the pre-nominal form has not continued its expected decline. Wedgwood (1995), discussing this issue, concludes that the two genitive orders are differentiated in Modern English on the basis of the animacy of the modifier. The distribution of the prenominal variant strongly favours animate modifiers whereas the Saxon genitive appears predominantly with inanimate modifiers:

(4.9) a. the man's face

b. ?? the clock's face

(4.10) a. ? the face of the man

b. the face of the clock

As Wedgwood (1995) points out, linguists vary in the marking of grammaticality of these sorts of examples (for example, Hawkins (1981) and Huddleston (1984) use '?' to suggest gradient acceptability, whereas Giorgi & Longobardi (1991) also use '*' for some sentences). The important point to make about examples such as these is that native speakers have an intuition about this acceptability, and the main determinant of their judgements seems to be the animacy of the modifier. This is not what we might expect given EIC. Instead, we should expect the prenominal genitive to be less acceptable as the *length* of the modifier increases. This is because, in a construction such as $_{PP}[P\ _{NP}[NP\ N]]$, the length of the CRD of PP increases with the length of the genitive NP. Although this is clearly a factor in determining the acceptability of genitive constructions (witness the ready acceptability of *the face of the friendly man next door*), it cannot predict the judgements given above and is incompatible with borderline cases such as:

(4.11) a. the friendly man's face

b. ? the face of the friendly man

So, animate modifiers are possible prenominally regardless of length, whereas post-nominal modifiers will tend to be either inanimate or animate and long. This could mean that there is some processing pressure acting to counter the EIC that prefers animates to be early in the utterance for some reason. We have no reason for believing this at the moment, however, and it is sensible to look for a simpler explanation that does not require us to posit any extra unmotivated psycholinguistic machinery.

Instead, we will simply say that the two types of genitive are functionally differentiated, with the prenominal type 'attracting' animate modifiers. The acceptability judgements given on the basis of animacy are thus not the result of some unknown functional pressure applied on the fly, but instead are coded for as part of native speaker competence. The form of this coding is debatable since it does not result in reliable grammaticality judgements — it may be that prenominal genitives with inanimate modifiers are produced as marked variants by analogy with a 'basic' animate variant. This functional differentiation (however it is coded for) stops the process of adaptation since it does not allow selection to operate on the two genitive orders. The remaining question is whether this differentiation is itself unpredictable or if it too can be related to the adaptive process operating with constraints set by the architecture of grammar.

4.5 Limits on grammatical primitives

The parsing theory discussed in chapter 2 relies on the size in numbers of words of constituent recognition domains. This makes sense in terms of parsing and may eventually be reducible to a theory of working memory, the idea being that the amount of information that has to be held in working memory and/or the time that it has to be held there for are directly related to the difficulty in accurately processing that information. This means that processing complexity is a gradient phenomenon which 'counts' numbers of words (c.f. Frazier (1985) who puts a discrete limit on the size of 'viewing window' in her parser, and Berwick & Weinberg (1984) whose parser also has an upper bound). This is markedly different from what we see in grammars, which seem unable to count words in this way. In other words the grammar is unable to *directly* reflect processing preferences.

As has been shown, however, there is overwhelming evidence that the grammars of the world's languages *have* responded to parsing. Instead of putting constraints on numbers of words, constraints are placed on positions of syntactic categories, each of which have different average numbers of words in texts. In this way, the architecture of grammar forces the acquirer to *reanalyse* patterns in the trigger in terms of category rather than length. So, if a prepositional language has prenominal modifiers, then the parser will filter many of the ModN constructions from the trigger. The likelihood of a construction being well represented in the trigger depends on the number of words in the modifier. The acquisition process cannot capture this generalisation, however; instead, the distribution of ModN constructions in the trigger is misanalysed as being dependent on the syntactic category of the modifier. Since relative clauses are likely to be longer than other modifiers, these are most likely to be barred from prenominal position (and so on down the hierarchy). So, all the examples given in this thesis so far have implicitly assumed some role for constraints on grammatical primitives, otherwise we would expect to find prepositional languages that preposed modifiers less than 3 words long and postposed others, for example.

4.5.1 Heavy NP shift

The sentences below demonstrate a weight (i.e. number of words) based "rule" that exemplifies the limits of grammaticalisation (from Rickford *et al.* 1995, see also §2.1.2):[8]

[8]These are traditionally Particle Shift examples. The terminology is unimportant, but I will subsume these under the term "Heavy NP Shift" because it makes the proposed motivation for the rearrangement

(4.12) a. *bring up it

b. bring it up

c. ??bring the subject we were talking about last night up

d. bring up the subject we were talking about last night

(4.12c) has a long NP interrupting the early processing of the MNCCs of the VP and is hence difficult to process, although we would hesitate to call it ungrammatical. The shifted example (4.12d) is much better in comparison. Notice that the shifted example (4.12a) is actually ungrammatical. The shift in this case from (4.12b) involves only one word, the pronoun *it*, and therefore brings no advantage in terms of parsing.

The grammatical situation suggested by these examples is quite complex. In response to pressure from parsing, there seems to be a grammatical variant ordering with NP shifted rightwards in the VP (the actual syntactic structure of the construction need not concern us here). The non-shifted variant is not ungrammatical since for many NPs it does not cause a serious problem for parsing. Therefore, both orders must be grammatical because the grammar cannot stipulate a certain number of words above which the NP is too long to stay before the particle. The grammar has responded to the case where the nominal is only a single word by making the postposing of *pronouns* ungrammatical. This is possible since the grammar can make a distinction between pronouns and full noun phrases. This is a case where a length-based distribution is reanalysed as a category difference, hence the grammaticality of (4.13a-b) even though the NP is also only a single word:

(4.13) a. bring up Fred

b. bring Fred up

All the above is received orthodoxy in linguistics and seems to fit well into the theory of selection constrained by the architecture of grammar. The idea that the occurrence of heavy NP shift in texts is determined by numbers of words in the NP has been challenged, however (see Rickford *et al.* 1995, 117 for a review). Notice, that if

clear.

we adopt the position of not assuming speaker altruism then we are only able to make predictions about the *acceptability* of the NP-shift sentences, not their distribution in texts. If the latter were determined by weight, then it would mean that speakers were responding to the needs of hearers in shifting NPs.[9]

Rickford *et al.* (1995) present a preliminary statistical study of heavy NP shift in texts and conclude that number of words is not the most significant determinant. Instead they point to a determinant based on the syntactic structure of the NP. According to their results, NPs with embedded sentences are more likely to be shifted, followed by conjoined NPs or NPs containing PPs. Simple NPs with or without modifiers are the least likely to be shifted. From our point of view, there are some problems with this analysis. In order to test the impact of EIC on production of NP shifted sentences we would need to know not only the number of words in the NP, but the number of words in the constituent shifted over. However, that aside, if the syntactic structure of the NP is important, it is interesting to speculate on whether there is a grammatical constraint on heavy NP shift after all. Again, it is not clear what form this constraint would take (especially considering it cannot be an exceptionless one), but if this is the case then it demonstrates another way for a grammatical rule to *approximate* to a length-based rule without actually referring to numbers of words.

4.5.2 Animacy and length

Returning to the problem posed by the English genitive construction, it is tempting to consider whether the fixing of the order of animate genitive modifiers prenominally is not also driven by processing. In a prepositional language, the parser prefers genitives to be postnominal, but if they do appear prenominally, then short genitive modifiers are preferred. The history of the English genitive for a certain period suggests that the pressure from the parser was resulting in the grammaticalisation of a postnominal order and the removal of the prenominal genitive. Given the distribution of orders in the trigger (with heavy prenominal genitives tending to be filtered out by the

[9] It has already been pointed out that the assumption of speaker altruism is not incompatible with the account put forward in this thesis, but it should not be the taken as the null hypothesis.

parser) this is the expected response of an acquisition device that is unable to express regularities in the input in terms of numbers of words. Modern English has not continued this trend. Is it possible that the modern language has 'discovered' another way of expressing the processing preference apparent in the trigger experience — one that differentiates between GenN and NGen in a way that reflects the relative lengths of the two types of modifier? This boils down to whether animates tend to be shorter than inanimates.

In order to answer this question Wedgwood (1995) looks at the lengths of relevant animate and inanimate genitive modifiers in a random sample from the LOB corpus of present-day English.

> "A 'relevant genitive' is here taken to be some attributive relation between two nouns using (in one sample) the *of*- and (in the other) the *-'s* constructions, which, except for the differentiation by animacy, is potentially expressible using either construction." (Wedgwood 1995:23)

The genitives are then split into two categories on the basis of the animacy of the modifier and their length distribution recorded. The results are reproduced below (Wedgwood 1995:24):

Length	1	2	3	4	5	6	7	8	9	10	>10
Animate	103	98	12	4	2	1	0	2	0	1	1
Inanimate	15	72	40	11	12	5	1	2	2	0	3

This distribution shows that, statistically, animates are significantly shorter than inanimates.

The picture that results from this is that at some point in the history of English, the parser-imposed distribution of genitive modifiers (short genitives prenominally) was reanalysed by language acquirers as reflecting a preference for animates prenominally. This is understandable since the longer modifiers that would be filtered out in prenominal position were more likely to be inanimate. This animacy distinction can be expressed by the grammar whereas length cannot, so the two types of genitive became functionally differentiated on the basis of animacy. Now it is impossible for selection to continue to work since the two genitives are no longer in competition in Kroch's (1994) terms. Interestingly enough, the selective process, constrained as it was by possible grammatical primitives, fails to result in the 'perfect' adaptation (only allowing NGen), and the possibility of long prenominal animate genitives is retained.

If the length difference between animates and inanimates could be shown to be universal, then we can make an interesting prediction about precedence rules in languages that make a grammatical distinction based on animacy. We should find at least in a significant number of cases that the order of animates versus inanimates follows the order of heads and modifiers in the language. Although such a study is beyond the scope of this work, it is interesting to note that Morolong & Hyman 1977 (cited in Hawkins 1994a, 424) describe a rule in Sesotho grammar which mirrors the English genitive case. If dative and patient noun phrases are both animate or both inanimate they may appear in either order in the Sesotho clause. Where they differ in animacy, the animate comes first. Given that Sesotho constituents are recognised on their left boundary, it is more efficient for long NPs to appear late to minimise recognition domains. This appears to be grammaticalised in terms of animacy. Further work needs to be done in this interesting area, particularly to check if head final languages with animacy-based rules tend to order animates late.

A final example involves German datives and accusatives. Consider the following sentences:

(4.14) a. Ich gab es ihm
I gave it(acc) him(dat)
"I gave it to him"

b. ? Ich gab ihm es
I gave him(dat) it(acc)

(4.15) a. ? Ich gab das Buch ihm
I gave the book(acc) him(dat)

b. Ich gab ihm das Buch
I gave him(dat) the book(acc)

c. ? Ich gab dem Mann es
I gave the man(dat) it(acc)

d. Ich gab es dem Mann
I gave it(acc) the man(dat)

(4.16) a. ? Ich gab das Buch dem Mann
I gave the book(acc) the man(dat)

b. Ich gab dem Mann das Buch
I gave the man(dat) the book(acc)

Sentences (4.14a-b) show that, where both dative and accusative are of minimal length (i.e. pronouns), there is an arbitrary grammaticalised ordering principle for accusative first.[10] However, when one of the nominals is a full NP and the other is a pronoun, parsing considerations have been grammaticalised so that the pronoun is strongly preferred before the NP (4.15a-d). Finally, where both nominals are full NPs, and hence could potentially vary in length considerably, the preferred order is dative before accusative. James Hurford (personal communication) has suggested that this may be because datives are typically more animate than accusatives and, as we have already shown, animates are on average shorter than inanimates. Hence, these examples show two ways in which a short-before-long parsing preference has been grammaticalised in German: firstly, based on the difference in prototypical lengths of pronouns vs. full NPs, and secondly based on the difference in prototypical animacy of dative and accusative, and derivatively, their lengths.

[10]Although, notice that there is the same problem here as with the English genitive examples regarding the judgement of ungrammaticality

4.6 Implications for linguistic theory

The discussion in this chapter has highlighted the importance of examining both processing considerations and formal models of syntax in explaining the origin of language universals. Both the parser and the innate language acquisition device leave their mark on language, but it is only by taking into consideration both mechanisms that the role of each can be uncovered. The diagram in figure 4.5 shows the different possible classes of language. E is the set of logically possible languages; L is the class of learnable languages, its boundary set by the innate language acquisition device; and F is the class of languages predicted to occur given the selection theory of chapters 2 and 3. Obviously, the languages we should expect to occur are those in $F \cap L$. Some of the languages predicted by the application of parallel function to the selection model do not occur because they are in the set $F \cap \overline{L}$. Similarly there may be languages that do not occur but are perfectly learnable in the set $\overline{F} \cap L$. These are ruled out by considerations of processing. I would argue that many of the language types that are barred by the universals considered in this thesis are in this set. So, for example, a language with oblique relatives but no direct object relatives is ruled out because of the interaction of p- and m-complexity in the Arena of Use. There is nothing that should lead us to believe that such a language is actually unlearnable.

This diagram fails to capture some of the more subtle interactions discussed here, however. We have seen that (a) languages *can* arise that respond to parallel function, albeit in unexpected ways, and (b) the animacy distinction in the English genitive is explicable in terms of processing, although the outcome does not fit into the general pattern of adaptation. The acquisition device in a sense provides ad-hoc solutions to the problem of representing in I-language the pressure exerted by processing on E-language. What these 'solutions' will be is fairly predictable, although sometimes the outcome is unexpected. In the English case there was a reanalysis of the underlying regularities in the input data — a length difference was reinterpreted using the grammatical primitive, animacy. Hopi, because of the idiosyncrasies of its morphology (resulting in an interaction of the switch-reference behaviour of the relative pronoun

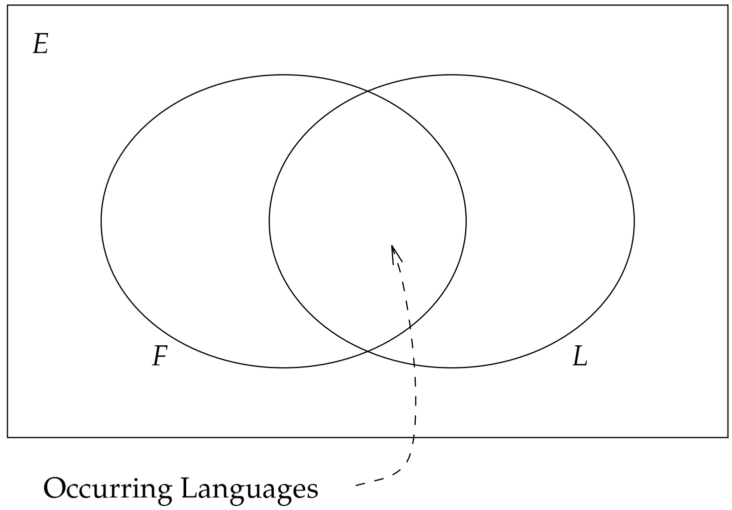


Figure 4.5. Interacting constraints on possible languages.

and subject case marking) has a mechanism for coding a constraint on the matrix function of object relatives. Can we say that this is an adaptation to the pressure exerted by the parser against S^O relatives? We cannot tell, although we might expect that there would be a pressure to change Hopi morphology if S^S relatives were made impossible.

The processing mechanisms make selections among utterances, and those selections cannot inform the acquisition device except by filtering input from the trigger. The resulting changes in the grammar of the language may lead to the removal of the particular structures that cause problems for processing, but they may not. If we are to gain a deeper understanding of the origins of universals we need to look for all the processing pressures that might be involved *and* what role the effect of those pressures on the trigger might play in the process of acquisition. The advantage of this approach is that troublesome counter-examples from the functional perspective may be mitigated by looking into constraints imposed by the architecture of grammar; from another perspective the burden of explaining all constraints on distribution uncovered by typology can be lifted from a theory of the structure of an innate UG.

We are now at a point where the functional and formal (or innatist) perspectives are mutually reinforcing rather than competing as they appear to be from so much of the literature (see Hurford 1990 for review). The recourse to innate/formal constraints might seem to raise more questions than it answers. For example, can the particulars of a formal theory of UG themselves be derived from other factors, or are they mere stipulations set up to account for the data? Recent developments in the literature suggest a way in which functional considerations may directly influence the structure of the innate language faculty. The next chapter considers this final thread in the web of function, selection and innateness.

Chapter 5

Innateness and function in linguistics

It is widely believed by linguists that the human ability to acquire language is at least in some part innately given, and that UG in the Chomskyan sense is embodied in this ability. Indeed, this assumption has been implicit in much of the discussion in this thesis so far. The previous chapter showed that such an innate LAD is required in combination with a theory of linguistic selection in order to fully understand the fit of universals to processing pressures. Recent research has begun to look at the possibility of examining the origins of the features of this innate faculty themselves, arguing that these too may have their roots in essentially functional pressures. This final chapter reviews some of this recent literature and examines whether it poses a competing theory of the origin of universals.

5.1 Natural selection and the LAD

Christiansen (1994) characterises explanations of the origin of an innate language acquisition device into two types: *exaptationist* and *adaptationist*. Proponents of the first type of explanation — among whom Christiansen cites Chomsky (1988)[1] and Piattelli-Palmerini (1989) — argue that natural selection plays only a minor role in the evolution of the complex domain-specific LAD. Instead they use the term *exaptation* (Gould & Vrba 1982) to describe the mechanism whereby the neural structures supporting language acquisition evolve. Exaptation refers to the reappropriation of form for some purpose other than the one that drove its evolution. Indeed Gould & Lewontin (1979) admit the possibility that the structure that is exapted may have no prior function at all, but simply be a "spandrel".[2] In this view then, the LAD might simply be a by-product of increased brain size, for example.

The adaptationist perspective (e.g. Pinker & Bloom 1990, Hurford 1989, Hurford 1991) places the burden of explaining the origin of the LAD on natural selection. In particular the LAD is claimed to have evolved through selection *for the function it now fulfils*. This relies on the assumption that human language confers a survival or reproductive advantage on the organisms that have it. This assumption seems to be fairly well accepted, although when we get to specific features of UG (see below) there seems to be greater unease. Lightfoot (1991:69), for example, pours scorn on the adaptationist argument suggesting "the Subjacency Condition has many virtues, but I am not sure that it could have increased the chances of having fruitful sex." We should reject Lightfoot's complaint because it relies on the "argument from personal incredulity" in Richard Dawkin's words. It rejects the adaptationist position simply because it is hard to believe, but where is the alternative? That the LAD evolved as an adaptation to acquisition should be our null-hypothesis — after all, natural selection is the most successful explanation of adapted complexity in nature that we have — so the burden is on the exaptationists to come up with an alternative explanation. (We

[1]The views of Chomsky on the evolution of language are notoriously difficult to unravel. In some papers he seems to suggest that the LAD can be viewed from an adaptationist perspective (e.g. Chomsky & Lasnik 1977, Chomsky 1980). A complete review of his views on this point would be a major undertaking, however see Newmeyer's (1994b) for an interesting perspective.

[2]The term "spandrel" is an architectural one, referring to a space formed at the meeting of two arches. At the San Marco basilica in Venice these spandrels are filled with a mosaic design which perfectly fits the triangular space provided. Gould and Lewontin point out that this apparent design should not lead us to believe that the function of the arch is to provide the artist with a space for a mosaic. Instead the spandrel is a by-product of the arch which has been adapted, or *exapted*, for an artistic function.

will return to the specific problem of subjacency later.)

Of course, before we appeal to the adaptationist approach, we need to know in what way the LAD is *adaptive*.

> "Do the cognitive mechanisms underlying language show signs of design for some function in the same way the anatomical structures of the eye show signs of design for the purpose of vision? What are the engineering demands on a system that must carry out such a function? And are the mechanisms of language tailored to meet those demands?" (Pinker & Bloom 1990:712)

To begin to answer these questions, and bolster support for the idea that the LAD is an adaptation, Pinker & Bloom (1990:713-714) list some design features of grammars such as: major and minor lexical categories, major phrasal categories, phrase structure rules, linear order rules, case affixes, verb affixes, auxiliaries, anaphoric elements, complementation, control, and wh-movement. They claim that these features of grammars — which from our innatist perspective are properties of the LAD — work together to make "communication of propositional structures" possible. Notice that Pinker and Bloom are not talking about the particular instantiations of these features in languages, but their existence as features of Language. So, for example, linear order and case affixes "distinguish among the argument positions that an entity assumes with respect to a predicate"(p.713), suggesting their presence in UG requires an adaptationist explanation. However, notice that the *particular* word orders or case affixes found in languages are not an issue for Pinker and Bloom.

The general features of UG appear to be one possible evolutionary solution to the problem of acquiring and representing a communicative system that allows the transmission of propositional structures. This adaptationist argument does not exclude a role for exaptation. Hurford & Kirby (1995) commenting on Wilkins & Wakefield (1995) suggest that a faculty for some form of *proto-language* (Bickerton 1990) was a primate exaptation from neural structures serving mental representation, but the human LAD has adapted from this precursor. In a sense, any exaptationist argument must include some degree of adaptation, since it is highly improbable that a complex structure evolved to fulfil some function can, by coincidence, also be used for some other purpose. The real issue is at what point in evolutionary history the LAD began to evolve in response to pressures imposed by the function it now fulfils. To put it another way, how much of the current LAD can we ascribe to natural selection for linguistic communication? Pinker and Bloom's argument suggests that at least some of the most basic features of UG are adaptations for communicative purposes.

5.2 Newmeyer on function

If we accept the idea that the origin of the LAD necessarily involves some degree of adaptation to the function it currently fulfils, and furthermore that the "basic design features" of Pinker & Bloom (1990) are the result of this adaptation, we are led to an interesting conclusion about more specific features of UG. Since the adaptation of the LAD to communicative ends must occur after any exaptation of neural structures, the more specific to language a mental feature is the more likely it is to be the result of an adaptation. The fact that, say, the presence of linear order rules are an adaptation to communicative ends suggests that the Subjacency Principle, for example, must also be viewed as an adaptation. To say otherwise would be to suggest that the Subjacency Principle is a left-over from some other neural function whereas the presence of linear ordering in language is not, yet no non-linguistic parallel of subjacency has been proposed but it is easy to think of non-linguistic domains in which linear ordering is important (in the formulation of plans, for example).

5.2.1 Autonomy

This common-sense argument raises the obvious challenge of explaining the specific architecture of a Chomskyan UG in terms of adaptation to the function of communication. Rather surprisingly given the repeated claims of Chomsky that UG is innate and the demonstrated success of neo-Darwinian explanations of biological complexity, this challenge has until recently been ignored. Part of the reason may be that adaptationist explanations appear to be at odds with the *assumption of the autonomy of syntax*, which states that "there exists a set of nonsemantic and nondiscourse-derived grammatical primitives whose principles of combination make no reference to system-external factors" (Newmeyer 1992, 783, see also Chomsky 1975). In other words, an autonomous syntactic component will make no use of information about external functional pressures nor will it include representations of those pressures.

This assumption, although allowing generative syntactic theory to progress rapidly has unfortunately caused linguists interested in functional explanation to generally reject generative syntax and some of those who accept the autonomy thesis to deny the possibility of functional explanation. This rejection of the link between function and autonomy is misguided in two ways.

Firstly, the simulations presented in this thesis explicitly take on board the assumption of autonomy in the design of the data structures that encode grammars. As discussed in chapter 2, they have purposefully been made as simple as possible: mere lists of possible utterance types. In no sense does the I-domain have any access to information about the processing complexity of the utterances they indirectly encode. Nevertheless the universals that emerge from the simulations clearly have a functional explanation. The end state of the simulation is that the particular distribution of grammars of the speech community collectively encode the processing pressures in the arena of use, without ever violating the autonomy of the individual grammatical knowledge of the language users. Furthermore, the evidence presented in chapter 4 *requires* the autonomy assumption for the functional explanation to work.

Secondly, as Newmeyer (1991) argues, functional considerations may directly shape the form of the syntactic component without violating its autonomy from function:

> "Despite the frequently voiced functionalist opinion that to identify a principle as innate is to abandon any attempt to explain it, there exists a well-accepted (functional) mechanism for explaining the provenance of innate traits: natural selection. It is quite plausible that the design of the grammatical model as a whole or *some particular grammatical principle* might

> have become encoded in our genes by virtue of its being successful in facilitating communication that the survival and reproductive possibilities of those possessing it were enhanced. In a sense, a functional explanation would hold at the evolutionary level.
>
> Thus autonomy is also compatible with a functional explanation for those aspects of language that form part of our biological endowment." (p. 7, emphasis my own)

Thus Newmeyer is going further than Pinker and Bloom in espousing a functional explanation for particular features of UG, rather than the broader design features of language. He also appears to admit the possibility that the same pressures that are appealed to by functional linguists can be applied to phylogenetic explanation.

5.2.2 Polystratal models of syntax and iconicity

The standard structure of the government-binding theory of syntax is shown in figure 5.1. The syntactic structure of a sentence is simultaneously represented at the various levels in the diagram which are related by a declarative transformational rule, move-α, whose role is to relate elements in particular positions at one level with the 'same' elements in different positions at neighbouring levels.[3]

This polystratal representation schema is part of the autonomous, innately given, architecture of grammar, but Newmeyer (1992) argues that it can be given a functional explanation in terms of *iconicity* (e.g. Haiman 1985). Givón (1985) suggests that a syntactic *form* is easier to process if it is in an iconic relation with its *content*, if "the code is maximally isomorphic to the experience"(p. 189). If this is true, then we can expect that grammatical representations will be arranged in such a way as to favour iconicity. There are, however, many ways in which a form can be iconic, reflecting the several dimensions of 'content'.

[3]Recent developments in generative syntax (e.g. Chomsky 1992, Marantz 1995) have suggested a revision to this model involving a more derivational approach to move-α, and only two levels, although it is possible that this could be given a declarative interpretation with multiple levels. We will not discuss this here, but Newmeyer's discussion is probably consistent with this variant of the model.

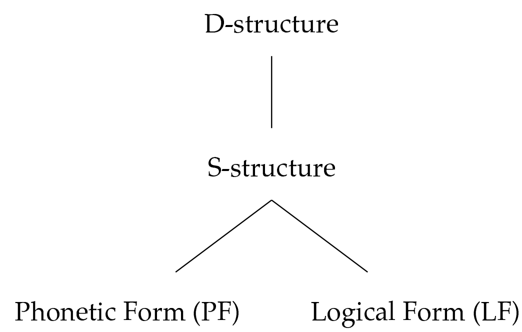


Figure 5.1. The polystratal architecture of the GB theory of syntax.

> "Maximal isomorphism for one property may not be maximal isomorphism for another. Consider, for example [Newmeyer's (32)]:
>
> (5.1) a. Who did Mary love?
>   b. Mary loved everyone.
>   c. Mary loved John.
>
> These three sentences have identical predicate-argument relations; their D-structure representations are thus identical, roughly as in [Newmeyer's (33)]:
>
> (5.2) a. Mary loved who.
>   b. Mary loved everyone.
>   c. Mary loved John.
>
> But at the level at which (5.1a-c) are represented identically, it is not easy to capture in any elegant way the fact that the quantification relations in (5.1a-b) differ profoundly from those in (5.1c), which is not an operator-bound variable construction semantically." (Newmeyer 1992:788–789)

So, we have a conflict here between an iconic representation of predicate-argument relations and quantifier-variable relations. Both cannot be represented in an iconic fashion at the same level. Instead, the former is represented at D-structure as above, and the later at LF as (Newmeyer 1992:788):

(5.3) a. Who$_i$ [Mary love e$_i$]

b. Everyone$_i$ [Mary love e$_i$]

c. Mary love John

The word order of utterances may not always reflect one or other of these levels partly because of considerations of processing such as heavy constituent shift. This motivates the presence of a the third level, S-structure (Newmeyer does not discuss PF).

Polystratal representations of syntactic structure as part of our biologically given faculty for language have arisen for functional reasons during the evolution of our species. The pressure for iconic representations — ultimately in response to processing needs — has favoured syntactic structures in which the 'same' elements (i.e. elements that are related by move-α) can enter into different iconic relations at different levels. In this way, Newmeyer approaches a basic assumption of autonomous syntax from a functional perspective.

5.2.3 Principles and processing

In another important paper Newmeyer (1991) goes further with the idea that processing can ultimately explain the nature of many of the specific principles of UG, also without compromising the autonomy thesis.

> "We have already seen that the model of autonomous grammar . . . has features that suggest it was shaped by natural selection, that is, that it evolved to its present state in effect *because* it was functionally so advantageous. It will be argued . . . that the same is true of the central principles of autonomous syntax. These principles were encoded in our genes by virtue of their playing such a central role in communication that the survival and reproductive possibilities of the species were advanced as a result of them." (p.12)

One of the examples that Newmeyer gives is Subjacency (Riemsdijk & Williams 1986, 62, cited in Newmeyer 1991, 12):

Subjacency condition No rule can relate X, Y in the structure

$$\ldots X \ldots [\alpha \ldots [\beta \ldots Y \ldots$$

or

$$\ldots Y \ldots]\beta \ldots]\alpha \ldots X \ldots$$

where α, β are bounding nodes.

In English, the bounding nodes are IP and NP, hence the ungrammaticality of a sentences below where *who* has moved over two bounding nodes (with no intermediate 'landing site'[4]):

(5.4) a. *I met the fan who$_i$ we played $_{NP}$[the song which$_j$ $_{IP}$[t_i liked t_j]]

b. *Who$_i$ did $_{IP}$[Matt tell you when$_j$ $_{IP}$[he had met $t_i t_j$]]

The standard assumption is that the subjacency condition is one of a set of constraints on the application of move-α that form part of our innate knowledge of language. Although there is some cross-linguistic variability in the inventory of bounding nodes, the constraint can, in principle, be applied to any language. How can the existence of this constraint be explained? Berwick & Weinberg (1984) point out that the subjacency condition tends to rule out sentences in which the distance between the wh-element and its co-indexed gap is long. As already discussed in chapter 4, there is a pressure from the parser to keep this distance to a minimum. Newmeyer's argument is that this parsing pressure led to the biological selection of a language acquisition device that had some way of eliminating the worst wh-extractions from the language. Crucially, the resultant constraint does not make any reference to parsability, or even distance, but is an autonomous principle which tends to rule out particularly long-distance movement.[5]

[4]See, for example, Haegeman 1991, §6.2 for further details of the applicability of the subjacency condition.

[5]We will review other perspectives on the subjacency condition later in this chapter.

Newmeyer (1991:13) goes on to suggest that Principle A of the binding theory and the Empty Category Principle, have similar functional motivations. They both constrain the syntactic positions of anaphoric elements and their antecedents, which suggests that they may also aid the parsing of co-indexed elements. Newmeyer, however, does not go into this parsing motivation in any detail, so these principles will not be discussed here. Suffice to say that both principles also do not make reference to "system-external factors" even though an explanation of their origin can be conceived in terms of parsing pressures.

5.3 The LAD and universals

The previous section sketched a view of functional explanation that is rather different from the one put forward in this thesis. Various design features of the LAD and innate principles appear to show the kind of evidence of fit that was introduced in chapter 1. But, this "appearance of design" is precisely what we observed in the universals of previous chapters. Is it possible, then, that the phylogenetic approach to explanation proposed by Newmeyer can be extended to cover the same universals that have been the focus of this thesis, for example the word order universals of chapter 2? This type of explanation would only be available to us if we assumed that the word order universals we have looked at resulted from some innate constraint. This is in contradiction to what has been assumed so far, amounting to changing the diagram 4.5 in chapter 4 so that the area $\overline{F} \cap L$ (non-functional, learnable languages) is reduced to $\emptyset$ (see figure 5.2).

As discussed in chapter 2, one of the universals that Hawkins's (1994a) theory attempts to explain is the tendency for languages to have a consistent positioning of head relative to non-heads across phrasal categories. How might this be accounted for in terms of innate UG? As Giorgi & Longobardi (1991) point out, the development of X-bar theory (e.g. Jackendoff 1977) allowed for this regularity to be expressed as a generalisation over phrase structure rules, so that the rules specifying the order of head and complement can be expressed as $X' \rightarrow X\ XP$ or $X' \rightarrow XP\ X$, with

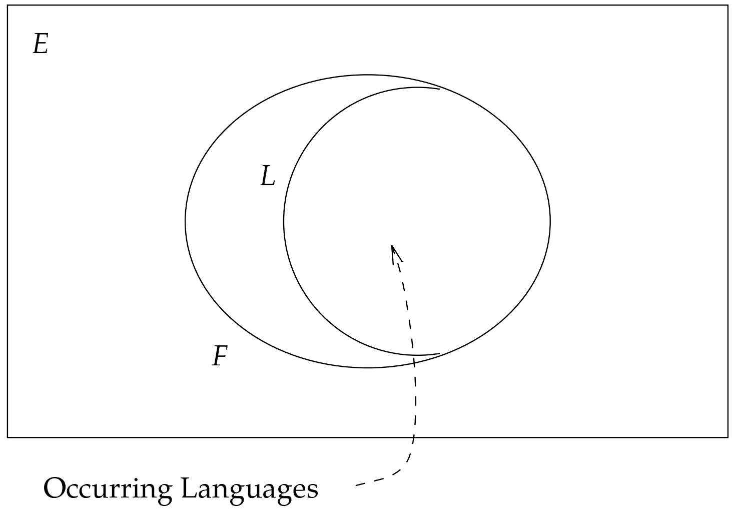


Figure 5.2. Possible languages where universals are explained exclusively by a functionally motivated LAD.

X ranging over the set of lexical categories. Later, after Stowell's (1981) rejection of phrase structure rules Chomsky (1986:88) simply refers to a "head-complement parameter" which can be either *head-final* or *head-initial* for a particular language.

Obviously, the problem with this approach to the universal is that there are exceptions to the generalisation — not all languages are consistently head-initial or head-final, although they *tend* to pattern that way. Travis (1984) looks at the word order of Modern Mandarin with respect to the head-complement parameter. She points out that NPs are head final, and certain PPs appear preverbally also suggesting that the parameter is set to head final. However, some PPs and direct object NPs can appear to the right of the verb. Furthermore, Modern Mandarin has prepositions rather than postpositions. For example (Travis 1984, 46, from Li & Thompson 1975, 180):

(5.5) a. ta gei wo mai le chezi le
he for me sell ASP car ASP
'He sold a car for me'

b. ta mei gei wo chezi le
he sell to me car ASP
'He sold a car to me'

(5.5a) is an example containing a pre-verbal benefactive PP, whereas (5.5b) contains a post-verbal dative PP. Travis (1984:48–53) argues at length that the difference between these types of PP can best be characterised as a difference in the assignment of the θ-role to *wo*. In the post-verbal case, she argues that the θ-role is assigned by the verb, whereas in the pre-verbal case the θ-role is assigned by the preposition. She then goes on to propose another parameter governing word order:

> "... the direction of θ-role assignment is another parameter which determines word order in languages. We can claim that while [Modern Mandarin] is head final, it assigns θ-roles to the right. If we look at the two categories that assign θ-roles, prepositions and verbs, we see that both of them appear to the left of the NPs to which they assign θ-roles. We will assume that within NPs, θ-roles are assigned by the preposition and not by the head N." (Travis 1984:53–54)

Only a flavour of Travis's account can be provided here, but she goes on to include another directional parameter: that of case-assignment. In this way different settings of the parameters can account for all possible orders of the two types of PP, and direct objects relative to the verb. This is because neither type of PP is case-marked by the verb, but the direct object is. So the case-assignment parameter may control the position of the direct object in relation to the verb independently of that of the PPs.

Where does that leave the observation that languages *tend* to pattern as head-initial or head-final? Firstly, notice that the head-ordering parameter can be in conflict with the other parameters. For Modern Mandarin, the head-ordering parameter defines the default ordering of constituents but the setting of the θ-marking parameter overrides this for the object and θ-marked PP. It could be argued, then, that all we need to account for the distribution of languages is for the contradictory setting of parameters to be marked in some way. Giorgi & Longobardi (1991:151) also argue that marked settings of parameters can account for cross-linguistic patterns, although they are looking at word order within the NP.

In summary, the innate LAD builds grammars with consistent head ordering as a default, but the setting of other parameters relating to the assignment of θ-roles and case may override these settings in the marked case. Stepping into Newmeyer's shoes, we might now say that the reason that UG is set up this way — that is, with default consistent head-ordering — is because of parsing. The EIC preferences for consistent ordering of MNCCs in this view influence the biological evolution of the LAD in order to constrain languages to aid parsing.

5.4 Biologisation or grammaticalisation?

Finally, we have come full circle: the cross-linguistic universals have been explained ultimately with reference to parsing. The problem is that there are now two candidate explanations for the same observed fit between universals and processing. A glossogenetic one in which languages themselves adapt to the pressures of transmission through the arena of use, and a phylogenetic one in which the LAD adapts to the pressures of survival in an environment where successful communication is advantageous. Looking at figure 5.3, we can see that if we accept Pinker & Bloom's (1990) approach, the difference between the functionalist and innatist positions is not in what explains language universals, but in the approach to solving the *problem of linkage*.

5.4.1 Subjacency five ways

To further highlight the lack of clarity in the literature regarding the connection between function, innateness and universals, we can return once again to the Subjacency Condition. At least five different positions are discernible on the issue of what this principle tells us about function and UG.

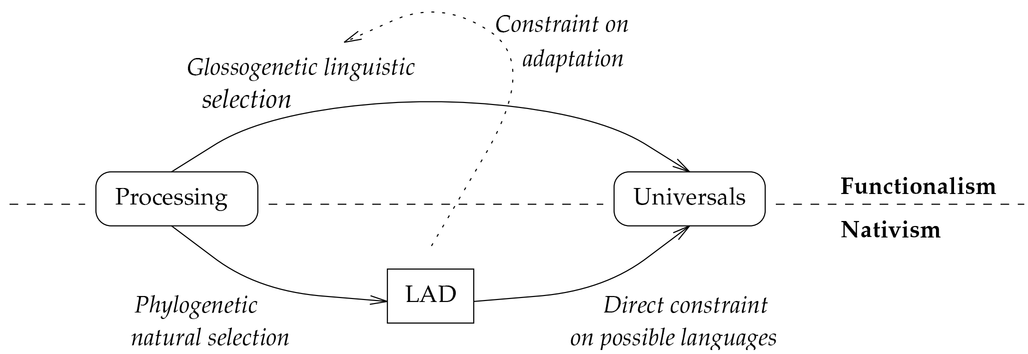


Figure 5.3. The (adaptive) innatist and functionalist approaches as solutions to the problem of linkage.

Piattelli-Palmerini 1989 As already discussed this author presents an exaptationist viewpoint on the emergence of the LAD. Part of the basis for his argument is the observation of *arbitrariness* in the formulation of UG principles (such as subjacency). The specific substance of the principle is not predictable as an adaptation to communication, therefore it lacks the appearance of design that is so typical of structures evolving through natural selection.

Pinker & Bloom 1990 In these authors' view, Piattelli-Palmerini's (1989) argument is flawed since there is nothing about evolution by natural selection that rules out arbitrariness. This is particularly true if communication is considered. The very nature of communication requires a shared coding protocol which may well be arbitrarily chosen from a set of equally functional options. Just because the specific principles that are innately coded cannot be predicted by looking at function, this does not mean that natural selection has not shaped those principles. Specifically, they argue that subjacency is an *arbitrary* compromise solution to pressures from expressiveness and parsing. "In the evolution of the language faculty, many 'arbitrary' constraints may have been selected simple because they defined parts of a standardised communicative code in the brains of a critical mass of speakers" (Pinker & Bloom 1990:718).

The subjacency condition could have been nativised in some other form, but to them the crucial point is that it must have been nativised somehow. In support of this, they cite Mayr (1982) on communication elsewhere in biology:

> "Behaviour that serves communication, for instance courtship behaviour, must be stereotyped in order not to be misunderstood. The genetic program controlling such behaviour must be "closed", that is, it must be reasonably resistant to any changes during the individual life cycle . . . " (p. 612)

Newmeyer 1991 This viewpoint has already been covered in section 5.2.3. It differs from Pinker and Bloom's mainly with regard to the importance placed on the parsability of subjacency violating structures. Newmeyer also stresses the pressure for evolution to constrain speakers of language in order to aid hearers — an issue which we will return to shortly.

Christiansen 1994 Whereas Newmeyer, and to a lesser extent Pinker and Bloom, use the heavy parsing complexity of subjacency-violating structures as evidence for the biological evolution of the constraint, Christiansen instead uses the same observation as evidence *against* an innate subjacency condition.

> "Since we therefore reasonably can construe subjacency simply as a constraint on processing . . . , it can no longer be considered to be an arbitrary linguistic phenomenon (as suggested by Pinker & Bloom 1990), but must indeed be conceived as a nonarbitrary byproduct of limited human processing abilities." (Christiansen 1994:130)

Notice that Christiansen appears to have missed the fact that Pinker and Bloom themselves appeal to the same evidence he does (i.e. the observations of Berwick & Weinberg 1984) to argue the opposite view.

Hawkins 1994a The final viewpoint on Subjacency is rather different from the others here since it rejects the existence of the condition altogether. Instead Hawkins proposes a wh-extraction hierarchy where each position on the hierarchy involves a movement spanning a larger structural domain than the positions higher on the hierarchy. Languages select positions on this hierarchy above which wh-extraction is grammatical, and below which it is not, in response to pressure from the parser.

Hawkins' argument against the classical interpretation of subjacency is based on a rejection of the "comp-to-comp" analysis of apparent violations of the condition. In this view, movements which appear to straddle two or more bounding nodes in fact take place in multiple stages, with the wh-element stopping off in intermediate positions (compare with 5.4b):

(5.6) Who$_i$ did $_{IP}$[Matt tell you $_{CP}$[t_i that $_{IP}$[he had met t_i]]

Here, the wh-element has moved from [Spec,CP] to [Spec,CP] and neither move violates the subjacency condition by crossing two IPs nodes. Hawkins (1994a) rejects this approach because of the lack of any independent psycholinguistic motivation for it. Notice, however, that it is just this kind of (partial) arbitrariness that other authors have used to argue for the innateness of the subjacency condition.

5.4.2 Speaker altruism again

At the moment it is a difficult task to choose between the five points of view summarised above, in the specific case of subjacency and in the general approaches to innateness and function that they suggest. The work presented in this thesis can shed light on some of the issues raised, however.

The evidence presented in chapter 4 should lead us to be wary of any approach that rejects an autonomous innate component altogether. In other words, there must be some *biologisation* of functional pressures involved, because the linguistic selection approach simply cannot explain the universals on its own. If this is the case we might wonder if there has been any glossogenetic adaptation at all.

One of the crucial features of Newmeyer's (1991) approach is his rejection of just this sort of glossogenetic functional explanation for language universals. He relies on an implicit rejection of speaker altruism in order to make his point:

> "In cases where ease for the speaker and the requirements of the hearer were in direct conflict, an obvious solution presented itself — to bypass directly the push-pull between speakers' demands and hearers' demands by incorporating those constraints necessary to the hearer directly into the innate language faculty itself. Thus the principles of UG were selected for, allowing a stable innate core to language, immune to the functional exigencies of the moment."(Newmeyer 1991:15)

In this way, Newmeyer rejects the possibility of particular languages evolving over a historical timescale to pressures from the parser. *If* speakers are not altruistic, he suggests, then there is no way in which hearers' needs could be reflected in grammars. And yet, subjacency (and indeed many of the universals we have discussed) appear to reflect just such one-sided needs. Hence, Newmeyer argues, they must have evolved phylogenetically.

Though there are certainly some innate constraints on acquisition that will ultimately be explained by appealing to functional asymmetries, it is a mistake to suggest that there will be a biological response wherever there is such a speaker/hearer difference. The simulations of chapters 2 and 3 show that languages may adapt glossogenetically to an asymmetric functional pressure, through a process of linguistic selection by the parser, even where there is not an innate constraint on them to do so. This weakens Newmeyer's argument considerably; linguistic selection and natural selection are both still, in principle, capable of explaining principles such as subjacency.

A more rewarding approach I would argue would be to admit the possibility of both kinds of adaptation and examine the mechanisms involved in more detail. It certainly seems likely given the quite different nature of the processes and objects that play a part in *biologisation* and *grammaticalisation*[6] that they will have observable differences once they are better understood. This thesis has gone some way to explore the glossogenetic adaptation and to provide a sufficiently general and explanatory

[6]This term is used in the sense introduced in chapter 2.

account of the universals examined in terms of linguistic selection. Inasmuch as this has been successful this should lead us to reject arguments that the universals thus explained, such as the tendency for consistent head-ordering, have an innate basis.

5.4.3 An approach to modelling the evolution of language

On the other hand, a similar investigation into modelling natural selection in language evolution might lead to an explanation of those universals that are more difficult for the linguistic selection approach. In this category we might put subjacency whose partially adaptive, partially arbitrary, characteristics are highlighted by Pinker & Bloom (1990), and also those constraints discussed in the previous chapter whose existence is predicted by 'failures' of glossogenetic adaptation.

What might such a model look like? More specifically, how would it counter Lightfoot's (1991) sceptical conclusion that the Subjacency Condition could not improve breeding success? A particularly promising line of work resurrects an evolutionary principle of Baldwin (1896), referred to as the Baldwin Effect. This effect predicts that a population of organisms that learn a beneficial behavioural pattern will tend, over time, to nativise that pattern. As French & Messinger (1994) note, the Baldwin Effect is still far from uncontroversial in biology possibly due to its apparent similarity to Lamarkian evolutionary principles, however in an important paper Hinton & Nowlan (1987) show that the Effect is completely compatible with neo-Darwinian assumptions.

Hinton and Nowlan examine the evolutionary dynamics of a population of organisms each with a set of 20 'neurons' which may be in one of two states. Each organism has a genotype that is made up of 20 genes, each of which has three possible alleles: 1, 0 or ?. The first two possibilities directly code for a corresponding neuron's state in that organism. The ? allele, on the other hand leaves the state of the corresponding neuron open to learning. The fitness of each organism (i.e. its chance of procreating) corresponds to the states of its neurons in such a way that for an organism to increase its fitness it must have *exactly the right set of neuronal states*. In other words, having only one neuron set wrong is just as bad as having all 20 wrong. The fitness landscape of this problem can envisaged as a flat (hyper-)plane with a single spike in it. Natural selection on its own has little chance of finding this spike, indeed it is no better than a random search at finding the 1 in 1048576 lucky individual with increased fitness.

The inclusion of the ? alleles, however, makes all the difference. In the Hinton and Nowlan simulation learning is implemented by an organism being allowed to try 1000 random settings of its neurons corresponding to ? alleles. If one of these attempts results in a correct setting of all 20 neurons (i.e. in combination with the neurons that are genetically specified), then this learning process stops. The chances of each organism being chosen as a parent in the creation of the next generation of organisms (which involves a simple recombination of genomes to create new individuals) depends on how quickly it reached the adaptive configuration.[7] Importantly, the learned settings of neurons are not passed on to the next generation (which would entail Lamarkian evolution), rather it is the sequence of 1,0 and ? that is used to form the offspring.

The original population of organisms each have on average 10 learnable neuronal settings, and 5 each of the pre-set 1s and 0s. During the simulation, the alleles specifying incorrect settings quickly disappear from the population, and the number of ? alleles decreases. As Hinton and Nowlan put it, learning has guided evolution. The reason for the relative success of nativising the correct settings when learning is involved, is due to the change in fitness landscape that the inclusion of ? alleles brings about. In the case where learning is not available, an organism near the correct combination is no fitter than one far away, but with the inclusion of learning, the landscape is *smoothed* so that an organism near the fitness peak is fitter (in the sense of quicker being able to get to the top) than one that is far away.

This effect has been elaborated in simulations by a number of researchers looking at various behaviours such food finding (e.g. Nolfi *et al.* 1994), carnivore avoidance (e.g. Ackley & Littman 1991) and even the evolution of (non-linguistic) communication

[7]The fitness μ of an individual $x_i \in (0\ 1\ ?)^L$ is a function of the number of learning attempts made g:

$$\mu(x_i(g)) = 1 + \frac{(L-1)(G-g)}{G}$$

where G is the maximum number of learning attempts allowed (here 1000). See Belew 1990 for an accessible analysis of the Hinton & Nowlan (1987) simulation.

(e.g. MacLennan 1991). In all these cases the ability of an organism to learn can guide evolution up to the peak of a fitness landscape. This highlights an attractive feature of the Baldwin Effect from our perspective. A common complaint regarding studies of the evolution of the human language faculty is that it is difficult to imagine a *gradual* evolution of the complex set of interacting constraints and principles that make up our language faculty. Such a faculty seems to us to be a "fitness spike" since without one component, how could the whole function at all? The simulations of the Baldwin Effect show that just such a structure can arise, however, as long as organisms have some ability to learn; in this way they can fill in the gaps in their innate ability with learned behaviour.

Turkel (1994) looks at a different aspect of the Baldwin Effect in order to explain the partially fixed, partially variable nature of UG. Assuming a principles and parameters model of this variation, he repeats Hinton & Nowlan's (1987) experiment and shows that a small shared set of variable parameters are the expected result of the learning guided evolution of language. The three alleles of Hinton and Nowlan correspond in this case to either invariant principles (0 or 1) or flexible parameters (?). Each parameter can be switched to 0 or 1 during learning, so the principles are assumed to be in some sense pre-wired parameter settings. Two organisms are potential communicators if their genomes *match*, where matching is possible if a 0 on one genome corresponds to a 0 or ? on the other, and similarly a 1 corresponds to a 1 or ?. Learning in the simulation involves randomly switching the parameters of each organism in a pair of potential communicators. The fitness of the organisms is related to the number of random settings it takes for both's sets of parameters to *exactly* match.

The result of Turkel's simulation is that the population converges on a set of shared principles and a small number of shared parameters. Which particular loci on the genome become fixed as principles, which remain as parameters, and whether the principles are set to 0 or 1 is completely arbitrary and different from one run of the simulation to another. The proportion of remaining parameters, however, shows little variation from run to run.

Another approach to modelling the evolution of language is presented by Batali

(1994) in an intriguing paper. Instead of relying on a rather abstract representation of principles and parameters as in Turkel (1994), Batali considers the possibility that a general learning mechanism can evolve to incorporate innate biases to particular classes of language that it is presented with. Specifically, he evolves a population of recurrent neural networks (e.g. Elman 1990) given the task of learning simple context free languages. Crucially, each network is given strings from a language with the same syntax, but with randomly chosen lexical items. The networks are thus unable to evolve to recognise exactly the language being presented. Instead, the generations of networks gradually improve in their ability to learn the languages they are presented with by nativising a disposition to learning *the particular class of languages in the simulation*.

The class of languages in the simulation can be described using a context free grammar (although the author does not present it in this way):

$$S \rightarrow Push\ M^*\ Pop$$
$$M \rightarrow Idle^*\ (S)\ Idle^*$$

So, each sentence in the language class starts with a $Push$, ends with a Pop, has any number of $Idle$s, and any number of other $Push$es and Pops as long as each $Push$ on the left has a corresponding Pop on the right. The individual languages differed in the assignment of four possible lexical items (a, b, c, d) to the three categories. So, for example, $baadcadcdd$ is a sentence in the language with the following assignment:

$$Push \rightarrow a$$
$$Push \rightarrow b$$
$$Pop \rightarrow d$$
$$Idle \rightarrow c$$

In order to parse a string in this class of languages, an automaton that knows the assignment of lexical items to categories must have some kind of counter. The counter will be incremented on encountering a $Push$ and decremented at each Pop. Each $Idle$ will not affect the counter. A valid string will return the counter to zero on encountering the last lexical item.

The networks in the simulation are each assigned a random language in this class and given the task of predicting when a sentence was finished (a good test of 'understanding' of the grammar without the need for supervised learning). The initial population of networks with random initial connections are fairly unsuccessful at this task after 500,000 characters of input. Selective breeding of networks on the basis of their final prediction ability is carried out so that the next generation has the *initial* connections of the best learners of the previous generation. Over (evolutionary) time, the performance of the networks improves markedly as the networks inherit an innate bias for learning this class of language. Specifically, the networks *learn* to associate $Push$ and Pop symbols with an internal counter, and have an *innate* association of the zero value of this counter with the end-of-string prediction.

Batali's work is particularly fascinating as it suggests a way in which to marry connectionist accounts of language learning with generative accounts of language acquisition. By modelling the evolution of general purpose learning machines, he has shown that there can be a gradual biologisation of the common features of the multiple learning tasks that face a population, leaving specific features to be learnt. Just as we saw in chapter 4 that language acquisition is a process of generalisation over input data, evolution here is generalising over learning problems. What remains to be explored is the extent of this kind of evolution's ability to generalise. If the distribution of input languages is constrained by functional pressures, what aspects of this distribution can the Baldwin Effect make innate?

Both Turkel's and Batali's simulations have their problems. For example, the particular settings of the innate principles in Turkel's evolutionary scenario are irrelevant to the fitness of the organisms — but how realistic is this? For Batali the most serious criticism could be that the actual languages that the networks learn are imposed by the experimenter rather than being generated by the organisms themselves, so how much can this tell us about the evolution of language? The value of these approaches, however, is in showing us that it is possible for natural selection to have shaped the human language faculty to *partially specify* the language we acquire. From looking at their results we can expect an innate LAD that evolved through natural selection to have some arbitrary constraints, but also to allow for variation.

In order for such models to solve the problem of the origin of specifically *functional* constraints as opposed to arbitrary ones, we would need to include functional pressures in the simulations. Batali's work shows that the broad design of the acquisition mechanism can become tailored to the problem of acquiring a class of languages, but this is far from the evolution of the particular constraints needed to rule out languages within this class that are harder to parse, for example. The Baldwin Effect shows us that gradual evolution of the LAD is possible, and that both arbitrary constraints and basic functional design features may become innate. It also shows us that there is a limit to this biologisation, since the models of its effect predict that the set of occurring languages will never be completely specified innately (see, e.g. French & Messinger 1994, Christiansen 1994, §5.2.2, and Belew 1990 for further discussion). The role of glossogenetic adaptation in the explanation of universals is therefore assured.

Chapter 6

Conclusion

Linguistic function in some broad sense interacts with linguistic form in two ways (although a third is possible):

1. Functional pressures influence natural selection which operates within physical and embryological constraints to give rise to an autonomous LAD, or Universal Grammar, over a biological timescale.

2. Functional pressures influence linguistic selection which operates within constraints imposed by Universal Grammar to give rise to observable language universals, over a historical timescale.

3. ?? The universals arising from linguistic selection may affect the biological evolution of the LAD through the operation of the Baldwin Effect.

One aim of this thesis has been to further illuminate (2) above. To do this new universals have *not* been uncovered, although some novel interpretations of the cross-linguistic data on case-coding have been proposed. Similarly, a new psycholinguistic model has *not* been proposed, although the separation of m- and p-complexity may be considered as a contribution to this area. Instead, the link between these two halves of the explanation — the process of linguistic selection — has been made explicit. By doing so, simulations can be designed that allow us to test the implicit assumptions of functional typology.

In its acknowledgement of the central role of the dynamics of language use and acquisition, this thesis places a good deal of importance on language change. Although the simulations deal with the behaviour of individual speakers, we have taken a *macroscopic* view of language change. In the study of universals we are essentially interested in the end result of all possible changes operating together; the relevant question being: is there a stable, emergent pattern cross-linguistically? In other words, individual changes in language and their causes are not our primary concern. It is however important that the model of change is not completely unrealistic or idealised. This is why it was considered important in chapter 2 that the behaviour of the simulation at a lower level matched the S-shaped curve observed by linguists interested in variation and change. In fact, to the best of my knowledge, this simulation is the first to *derive* the logistic curve that Kroch (1989a) imposes on the data. (Although, the recent work of Niyogi & Berwick (1995) mentioned in chapter 2, independently reaches a similar result.)

Another important result from chapter 2 is the conclusion that the assumption of speaker altruism is *not* required in order to explain the fit of universals to parsing pressures. Instead, this fit is the inevitable result of the parser having a selective influence on the transmission of forms through the arena of use. This is important since the assumption of speaker altruism is rather undesirable in the light of models of production, which rely on the modularity of the *conceptualiser* and *formulator* (Levelt 1989). Furthermore, it means that Newmeyer's (1991) innatist explanation is not the only possible one for the origin of universals that correspond to parsing pressures.

Chapter 3 poses the most serious challenge to any functional/typological view that simply assumes the link between processing and universals. The simulations show that hierarchical, or implicational, universals relating to relative clauses *only* emerge given competing functional pressures whose relative importance shifts over time. The stable, hierarchical universal is thus the result of a complex, unstable push-pull between speaker and hearer (although this competition may even be played out "within" one individual). A "type-graph" formalism suggested by Greenberg (1978) is used to help understand this result, although here it is clear that the simulation method itself is invaluable in testing the behaviour of the complex-adaptive systems model. The separation of two competing types of complexity in this chapter also suggests a reassessment of the case-coding distinction for relative clauses. The skewing, cross-linguistically, of various types of relative clause on the hierarchy is predicted on the basis of the relative morphological complexity of the strategy for forming each type. This seems to fit the available data rather well, although a larger scale typological survey is required.

Up to this point, the type of explanation examined relies solely on features of the arena of use (i.e. processing operating to select variant forms). However, an important finding of this thesis is that this type of functional explanation is incomplete without a consideration of the role of innate constraints on variation. This is demonstrated in chapter 4 where some features of innate UG act to limit and affect the adaptive process in interesting ways. It is only with a careful examination of these 'environmental' constraints imposed by our innate faculty that functionalist explanations can be saved from explanatory inadequacy (e.g. in the link between processing and relative clauses). It also helps us understand puzzling features of individual languages (such as animacy effects) as having their roots in apparently unrelated processing pressures.

To some the marriage of the functionalist approach and Chomskyan nativism may seem inappropriate. The assumption of the autonomy of syntax is at the core of the generative program and admitting language processing as a factor in the origin of linguistic structure appears to undermine this assumption. In chapter 5, this belief is attacked on two levels. Firstly, it is clear that the simulations of variation and change put forward here are quite compatible with the autonomy thesis. Secondly, a review of some of the recent literature on evolution admits the possibility of a functional underpinning for the autonomous syntactic principles themselves. Perhaps because such research is still at a preliminary stage, this chapter has raised many unanswered questions. We are left with a rather confusing picture of the multiple interactions of function, innateness and selection. After examining these interactions in terms of the link between processing and universals, however, I believe we can now at least ask the *right* questions.

Finally, the most important message of this thesis is that the problem of explaining universals goes to the very heart of most areas of modern linguistics. If we are to understand these emergent properties of language we need a more eclectic approach than is apparent in much of the literature. Whilst researchers dogmatically place themselves in the "functionalist" camp or the "formalist" camp we can only hope to see half of the picture.

Bibliography

ACKLEY, DAVID, & MICHAEL LITTMAN. 1991. Interactions between learning and evolution. In *Artificial Life II*, ed. by C.G. Langton, C. Taylor, J.D. Farmer, & S. Ramussen, 487–509. Addison-Wesley.

ANDERSEN, HENNING. 1972. Diphthongization. *Language* 48.11–50.

BAKKER, DIK. 1994. *Formal and Computational Aspects of Functional Grammar and Language Typology*. Dordrecht.

BALDWIN, J.M. 1896. A new factor in evolution. *American Naturalist* 30.441–451.

BATALI, JOHN. 1994. Innate biases and critical period's: Combining evolution and learning in the acquisition of syntax. In *Artificial Life IV*, ed. by Rodney Brooks & Pattie Maes, 160–171. MIT Press.

BELEW, RICHARD. 1990. Evolution, learning, and culture: computational metaphors for adaptive algorithms. *Complex Systems* 4.11–49.

BERG, THOMAS. 1995. *Language Processing as a Reason for Linguistic Structure and Change*. Manuscript.

BERWICK, R.C., & A.S. WEINBERG. 1984. *The Grammatical Basis of Linguistic Performance: Language Use and Acquisition*. MIT Press.

BICKERTON, DEREK. 1990. *Language and Species*. University of Chicago Press.

BYBEE, JOAN. 1985. *Morphology: A Study in the Relation between Meaning and Form*. John Benjamins.

——. 1988. The diachronic dimension in explanation. In *Explaining Language Universals*, ed. by John A. Hawkins. Blackwell.

CANN, RONNIE, & MARY TAIT, 1990. Free relatives revisited. Manuscript, University of Edinburgh.

CHOMSKY, NOAM. 1975. Questions of form and interpretation. *Linguistic Analysis* 1.75–109.

——. 1980. *Rules and Representations*. Columbia University Press.

——. 1981. *Lectures on Government and Binding*. Foris.

——. 1986. *Knowledge of Language*. Praeger.

——. 1988. *Language and Problems of Knowledge: The Managua Lectures*. MIT Press.

——, 1992. A minimalist program for linguistic theory. MIT Occasional Papers in Linguistics, No. 1.

——, & H. LASNIK. 1977. Filters and control. *Linguistic Inquiry* 8.425–504.

CHRISTIANSEN, MORTEN, 1994. *Infinite Languages, Finite Minds: Connectionism, Learning and Linguistic Structure*. University of Edinburgh dissertation.

CLANCY, PATRICIA, HYEONIJIN LEE, & MYEONG-HAN ZOH. 1986. Processing strategies in the acquisition of relative clauses: Universal principles and language-specific realizations. *Cognition* 24.225–262.

CLARK, ROBIN, & IAN ROBERTS. 1993. A computational model of language learnability and language change. *Linguistic Inquiry* 24.299–345.

COMRIE, BERNARD. 1981. *Language Universals and Linguistic Typology*. Basil Blackwell.

——, & EDWARD KEENAN. 1979. Noun phrase accessibility revisited. *Language* 55.649–664.

CORBETT, GREVILLE. 1983. *Hierarchies, targets and controllers: agreement patterns in Slavic*. London: Croom Helm.

CROFT, WILLIAM. 1990. *Typology and universals*. Cambridge: Cambridge University Press.

——. 1993. Functional-typological theory in its historical and intellectual context. *Sprachtypologie und Universalienforschung* 46.21–22.

CUTLER, A., J. HAWKINS, & G. GILLIGAN. 1985. The suffixing preference: a processing explanation. *Linguistics* 23.723–758.

CZIKO, GARY. 1995. *Without Miracles: Universal Selection Theory and the Second Darwinian Revolution*. MIT Press.

DAWKINS, RICHARD. 1982. *The Extended Phenotype*. Freeman.

DEVILLIERS, J. G., H. B. TAGER-FLUSBERG, K. HAKUTA, & M. COHEN. 1979. Children's comprehension of relative clauses. *Journal of Psycholinguistic Research* 8.499–518.

DRYER, MATTHEW. 1980. The positional tendencies of sentential noun phrases in universal grammar. *Canadian Journal of Linguistics* 25.123–195.

——. 1991. SVO languages and the OV:VO typology. *Journal of Linguistics* 27.443–482.

——. 1992. The Greenbergian word order correlations. *Language* 68.81–138.

DUBOIS, JOHN. 1987. The discourse basis of ergativity. *Language* 64.805–855.

ELMAN, JEFFREY. 1990. Finding structure in time. *Cognitive Science* 14.179–211.

——. 1991. Incremental learning, or the importance of starting small. In *Program of the 13th Annual Conference of the Cognitive Science Society*, 443–448, Hillsdale, NJ. Cognitive Science Society, Lawrence Erlbaum.

FAY, D. 1980. Transformational errors. In *Errors in Linguistic Performance: Slips of the tongue, ear, pen and hand*, ed. by V.A. Fromkin. Academic Press.

FISCHER, O. 1992. Syntax. In *The Cambridge History of the English Language, Vol II: 1066–1476*, ed. by N. Blake. Cambridge University Press.

FODOR, JERRY A. 1983. *The Modularity of Mind*. MIT Press.

FRAZIER, L. 1985. Syntactic complexity. In *Natural Language Parsing: Psychological, Computational, and Theoretical Perspectives*, ed. by D. Dowty, L. Karttunen, & A. Zwicky. Cambridge University Press.

——, & K. RAYNER. 1988. Parameterizing the language processing system: left vs. right branching within and across languages. In *Explaining Language Universals*, ed. by J.A. Hawkins. Blackwell.

FRENCH, ROBERT, & ADAM MESSINGER. 1994. Genes, phenes and the Baldwin effect: learning and evolution in a simulated population. In *Artificial Life IV*, ed. by Rodney Brooks & Pattie Maes, 277–282. MIT Press.

GELL-MANN, MURRAY. 1992. Complexity and complex adaptive systems. In *The Evolution of Human Languages*, ed. by J.A. Hawkins & M. Gell-Mann. Addison-Wesley.

GIORGI, ALASSANDRA, & GIUSEPPE LONGOBARDI. 1991. *The syntax of noun phrases: configuration, parameters and empty categories*. Cambridge University Press.

GIVÓN, TALMY. 1979. *On Understanding Grammar*. Academic Press.

——. 1985. Iconicity, isomorphism and non-arbitrary coding in syntax. In *Natural syntax*, ed. by John Haiman. Cambridge University Press.

GOULD, STEPHEN J. 1983. *Hens' Teeth and Horses' Toes*. Penguin.

——, & R.C. LEWONTIN. 1979. The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proceedings of the Royal Society of London* 205.281–288.

——, & E.S. VRBA. 1982. Exaptation — a missing term in the science of form. *Paleobiology* 8.4–15.

GREENBERG, JOSEPH. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In *Universals of Language*, ed. by Joseph Greenberg, 73–113. MIT Press.

——. 1966. *Language universals with special reference to feature hierarchies*. The Hague: Mouton.

——. 1978. Diachrony, synchrony, and language universals. In *Universals of Human Language 1, Method and Theory*, ed. by Joseph Greenberg, 61–91. Stanford.

GRIMM, H., H. SCHOLER, & M. WINTERMANTEL. 1975. *Zur Entwicklung sprachlicher Strukturformen bei Kindern*. Beltz.

GRIMSHAW, J., 1993. Minimal projection, heads and optimality. To appear in Linguistic Inquiry.

GROOS, A., & H. VAN RIEMSDIJK. 1979. Matching effects in free relatives: a parameter of core grammar. In *Theory of Markedness in Generative Grammar*, ed. by A. Belletti, L. Brandi, & L. Rizzi. Scuola Normale Superiore.

HAEGEMAN, LILIANE. 1991. *Introduction to Government and Binding Theory*. Blackwell.

HAIMAN, JOHN. 1985. *Natural Syntax: Iconicity and Erosion*. Cambridge University Press.

HALE, KENNETH, LAVERNE MASAYESVA JEANNE, & PAUL PLATERO. 1977. Three cases of overgeneration. In *Formal Syntax*, ed. by Peter W. Cullicover, Thomas Wasow, & Adrian Akmajian. Academic Press.

HALL, CHRISTOPHER. 1988. Integrating diachronic and processing principles in explaining the suffixing preference. In *Explaining Language Universals*, ed. by John A. Hawkins. Blackwell.

——. 1992. *Morphology and Mind: a Unified Approach to Explanation in Linguistics*. Routledge.

HAWKINS, JOHN A. 1983. *Word Order Universals*. Academic Press.

—— 1988. Explaining language universals. In *Explaining Language Universals*, ed. by John A. Hawkins. Basil Blackwell.

—— 1990. A parsing theory of word order universals. *Linguistic Inquiry* 21.223–261.

—— 1992a. Heads, parsing, and word order universals. In *Heads in grammatical theory*, ed. by Greville G. Corbett, Norman M. Fraser, & Scott McGlashan. Cambridge University Press.

—— 1992b. Innateness and function in language universals. In *The Evolution of Human Languages*, ed. by J.A. Hawkins & M. Gell-Mann. Addison-Wesley.

—— 1992c. Syntactic weight versus information structure in word order variation. In *Special Issue No. 4*, ed. by J. Jacobs, Informationsstruktur und Grammatik, Linguistische Berichte, 196–219. .

—— 1994a. *A performance theory of order and constituency*. Cambridge University Press.

——, 1994b. Some issues in a performance theory of word order. Manuscript, University of Southern California.

HAWKINS, R. 1981. Towards an account of the possessive constructions *np's n* and *the n of np*. *Journal of Linguistics* 17.247–269.

HEINE, BERND, ULRIKE CLAUDI, & FRIEDERIKE HUNNEMEYER. 1991. *Grammaticalization: a Conceptual Framework*. The University of Chicago Press.

HINTON, G., & S. NOWLAN. 1987. How learning can guide evolution. *Complex Systems* 1.495–502.

HOEKSTRA, TEUN, & JAN G. KOOIJ. 1988. The innateness hypothesis. In *Explaining Language Universals*, ed. by John A. Hawkins. Blackwell.

HUDDLESTON, R. 1984. *Introduction to the grammar of English*. Cambridge University Press.

HURFORD, JAMES. 1987. *Language and Number: the Emergence of a Cognitive System*. Cambridge, MA: Basil Blackwell.

——. 1989. Biological evolution of the Saussurean sign as a component of the language acquisition device. *Lingua* 77.187–222.

——. 1990. Nativist and functional explanations in language acquisition. In *Logical Issues in Language Acquisition*, ed. by I.M. Roca, 85–136. Foris Publications.

——. 1991. The evolution of the critical period for language acquisition. *Cognition* 40.159–201.

——, & SIMON KIRBY. 1995. Neural preconditions for proto-language. *Behavioural and Brain Sciences* 18.193–194.

HYMAN, L. 1984. Form and substance in language universals. In *Explanations for Language Universals*, ed. by B. Butterworth, B. Comrie, & O. Dahl, 67–85. Mouton.

JACKENDOFF, R. 1977. *X-Syntax: A Study of Phrase Structure*. MIT Press.

KAIL, M. 1975. Etude génétique de la reproduction de phrases relatives: 1. reproduction immédiate. *L'Année psychologique* 75.109–126.

KEENAN, EDWARD. 1972a. The logical status of deep structures. In *Proceedings of the Eleventh International Congress of Linguists*, ed. by L. Heilmann. Societa editrice il Mulino.

——. 1972b. Relative clause formation in Malagasy. In *The Chicago Which Hunt*, ed. by Paul Peranteau, Judith Levi, & Gloria Phares, 169–189. Chicago Linguistics Society.

——, & BERNARD COMRIE. 1977. Noun phrase accessibility and universal grammar. *Linguistic Inquiry* 8.63–99.

——, & BERNARD COMRIE. 1979. Data on the noun phrase accessibility hierarchy. *Language* 55.333–351.

——, & SARAH HAWKINS. 1987. The psychological validity of the accessibility hierarchy. In *Universal Grammar: 15 essays*, ed. by Edward Keenan, 60–85. London: Croom Helm.

KELLER, RUDI. 1994. *On language change: the invisible hand in language*. Routledge.

KIRBY, SIMON. 1994a. Adaptive explanations for language universals: a model of Hawkins' performance theory. *Sprachtypologie und Universalienforschung* 47.186–210.

——. 1994b. Constraints on constraints, or the limits of functional adaptation. In *Proceedings of the Edinburgh Linguistics Department Conference '94*, ed. by Simon Kirby, 48–54.

——. 1994c. A selection model explanation of the distribution of relativisation strategies on the accessibility hierarchy. In *Proceedings of the third Manchester postgraduate linguistics conference*. Forthcoming.

——. 1996. Competing motivations and emergence: explaining implicational hierarchies. *Language Typology* . Forthcoming.

KROCH, ANTHONY. 1989a. Function and grammar in the history of English. In *Language Change and Variation*, ed. by Ralph Fasold & Deborah Schiffrin, 133–172. .

——. 1989b. Reflexes of grammar in patterns of language change. *Language Variation and Change* 1.199–244.

——. 1994. Morphosyntactic variation. In *Papers from the 30th Regional Meeting of the Chicago Linguistics Society*, ed. by K. Beals.

LABOV, W. 1972. *Sociolinguistic patterns*. University of Pennsylvania Press.

LASS, ROGER. 1980. *On Explaining Language Change*. Cambridge University Press.

LEVELT, WILLEM. 1983. Monitoring and self-repair in speech. *Cognition* 14.41–104.

——. 1989. *Speaking: from Intention to Articulation*. MIT Press.

LEWONTIN, R. 1974. *The Genetic Basis of Evolutionary Change*. Columbia University Press.

LI, C., & S. THOMPSON. 1975. The semantic function of word order: A case study in Mandarin. In *Word Order and Word Order Change*, ed. by C. Li.

LIGHTFOOT, DAVID. 1989. The child's trigger experience: Degree-0 learnability. *Behavioural and Brain Sciences* 12.321–334.

——. 1991. Subjacency and sex. *Language and Communication* 11.3–28.

LUCE, PAUL A. 1986. A computational analysis of uniqueness points in auditory word recognition. *Perception and Psychophysics* 39.155–158.

MACLENNAN, BRUCE. 1991. Synthetic ethology: an approach to the study of communication. In *Artificial Life II*, ed. by C.G. Langton, C. Taylor, J.D. Farmer, & S. Ramussen, 631–657. Addison-Wesley.

MACWHINNEY, B. 1982. Basic syntactic processes. In *Language Acquisition: vol. 1, Syntax and Semantics*, ed. by S. Kuczaj. Lawrence Erlbaum.

——, & C. PLEH. 1988. The processing of restrictive relative clauses in Hungarian. *Cognition* 29.95–141.

MANZINI, R., & K. WEXLER. 1987. Paramaters, binding theory and learnability. *Linguistic Inquiry* 18.413–444.

MARANTZ, ALEC. 1995. The minimalist program. In *Government and Binding Theory and the Minimalist Program*, ed. by Gert Webelhuth. Blackwell.

MARCUS, MITCHELL P. 1980. *A Theory of Syntactic Recognition for Natural Language*. MIT Press.

MAXWELL, DANIEL. 1979. Strategies of relativization and NP accessibility. *Language* 55.352–371.

MAYR, E. 1982. *The Growth of Biological Thought*. Harvard University Press.

MCGILL, STEVEN, 1993. Linguistic evolution: language change, variation and selection. Manuscript, University of Edinburgh.

MOROLONG, M., & L.H. HYMAN. 1977. Animacy, objects and clitics in sesotho. *Studies in African Linguistics* 8.199–218.

NEWMEYER, FREDERICK J. 1991. Functional explanation in linguistics and the origins of language. *Language and Communication* 11.3–28.

—— 1992. Iconicity and generative grammar. *Language* 68.756–796.

—— 1994a. Competing motivations and synchronic analysis. *Sprachtypologie und Universalienforschung* 47.67–77.

—— 1994b. A note on Chomsky on form and function. *Journal of Linguistics* 30.245–251.

NIYOGI, PARTHA, & ROBERT BERWICK. 1995. The logical problem of language change. Technical Report AI Memo 1516 / CBCL Paper 115, MIT AI Laboratory and Center for Biological and Computational Learning, Department of Brain and Cognitive Sciences.

NOLFI, STEFANO, JEFFREY ELMAN, & DOMENICO PARISI. 1994. Learning and evolution in neural networks. *Adaptive Behaviour* 3.5–28.

PANOV, M.V. (ed.) 1968. *Russkij jazyk i socetskoe obščestvo, III, Morfologija i sintaksis sovremennogo russkogo jazyka*. Moscow, Nauka.

PIATTELLI-PALMERINI, M. 1989. Evolution, selection and cognition: from "learning" to parameter setting in biology and the study of language. *Cognition* 31.1–44.

PINKER, STEVEN, & PAUL BLOOM. 1990. Natural language and natural selection. *Behavioral and Brain Sciences* 13.707–784.

POLLARD, CARL, & IVAN SAG. 1994. *Head-driven Phrase Structure Grammar*. University of Chicago Press.

RICKFORD, JOHN, THOMAS WASOW, NORMA MENDOZA-DENTON, & JULI ESPINOZA. 1995. Syntactic variation and change in progress: Loss of verbal coda in topic-restricting *as far as* constructions. *Language* 71.102–131.

RIEMSDIJK, H. VAN, & E. WILLIAMS. 1986. *Introduction to the Theory of Grammar*. MIT Press.

SHELDON, AMY. 1974. On the role of parallel function in the acquisition of relative clauses in English. *Journal of Verbal Learning and Verbal Behaviour* 13.272–281.

——. 1977. The acquisition of relative clauses in French and English. Implications for language-learning universals. In *Current Themes in Linguistics*, ed. by F. Eckman. Hemisphere.

SOBER, ELLIOT. 1984. *The Nature of Selection: Evolutionary Theory in Philosophical Focus*. MIT Press.

SPERBER, D., & D. WILSON. 1986. *Relevance: Communication and Cognition*. Basil Blackwell.

SPIESS, ELIOT. 1989. *Genes in Populations*. John Wiley and Sons.

STOWELL, T., 1981. *Origins of Phrase Structure*. MIT dissertation.

TALLERMAN, MAGGIE. 1990. Relativisation strategies: NP accessibility in Welsh. *Journal of Linguistics* 26.291–314.

TAVAKOLIAN, S. 1981. The conjoined-clause analysis of relative clauses. In *Language Acquisition and Linguistic Theory*, ed. by S. Tavakolian, 167–187. MIT Press.

TIERSMA, PETER. 1982. Local and general markedness. *Language* 58.832–849.

TRAUGOTT, ELIZABETH C., & BERND HEINE. 1991. *Approaches to Grammaticalization Vol. I*. John Benjamins.

TRAVIS, L., 1984. *Parameters and Effects of Word Order Variation*. MIT dissertation.

TURKEL, WILLIAM, 1994. The learning-guided evolution of natural language. Manuscript, University of British Columbia.

VANEK, A.L. 1977. Aspects of subject-verb agreement. *Current Inquiry into Language and Linguistics* 23.

WEDGWOOD, DANIEL, 1995. Grammaticalisation by re-analysis in an adaptive model of language change: a case study of the English genitive constructions. Master's thesis, University of Edinburgh.

WILKINS, WENDY, & JENNIE WAKEFIELD. 1995. Brain evolution and neurolinguistic preconditions. *Behavioral and Brain Sciences* 18.161–226.

WITKOWSKI, STANLEY R., & CECIL H. BROWN. 1983. Marking-reversals and cultural importance. *Language* 59.569–582.