

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 organized bundles whenever life intervenes. The examples of this improbable order extend to the artefacts of life as well as to 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 artefacts. All these things are improbable in the sense that they inhabit a small, organized area in the much larger space of logical possibility.

This book looks at another phenomenon in the natural world: human language. The range of variation among languages is constrained in various interesting ways. ‘Language universals’ are statements which describe these constraints on variation. 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 artefacts.

The origin of this order is in itself interesting, but I shall be exploring a particular aspect of these constraints on variation which are shared by others in the natural world. This aspect can be termed ‘fit’ or ‘the appearance of design’. For example, trees appear to 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 efficiently to 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

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

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’.

Constraints on variation

I have mentioned that language universals can be thought of as mapping the constraints of variation on occurring languages in some more general space of logically possible languages. Figure 1.1 demonstrates how this works. In Figure 1.1*a*, the occurring languages (shown in grey) are evenly spread throughout the space of all logically possible languages (the whole space, labelled ‘E’, as is the convention with Venn diagrams). If this were the case, then there would be very little of interest that we could say about constraints on variation. In some sense, there would be nothing *cross-linguistic* to explain about languages. All we could hope to do is explain how *one particular language* happened to come to be in a particular point in the space ‘E’.

However, as previously mentioned, languages do not vary randomly—they do not evenly fill the space of possibility. In other words, the real situation is more like that in Figure 1.1*b*. Here, the languages cluster in a tightly constrained sub-part of the space. The boundary of this grey cluster is completely specified by the set of language universals.

What this cluster will look like and where it will be in diagrams such as these will depend on how we organize the space of logically possible languages. It is impossible to draw this space on a flat piece of paper, since it has many dimensions. Instead, typologists choose to look at a small number of dimensions when discussing a particular universal and consider how languages cluster when we draw the space ‘E’ highlighting only those dimensions.

Figure 1.1*c* shows a very simple example. Here, the space of logically possible languages has been arranged so that it is divided neatly into two types of language: those in which overt subject noun-phrases are possible, and those in which overt subject noun-phrases are impossible. All other dimensions along which languages may vary are ignored. It turns out that all languages allow overt subject noun phrases, so the grey cluster fills the left-hand side of the diagram.

This representation of the space ‘E’ is enriched in Figure 1.1*d*, where a second orthogonal dimension cuts across the first one. In this diagram, languages at the top of the space are those that do not possess vowels, and

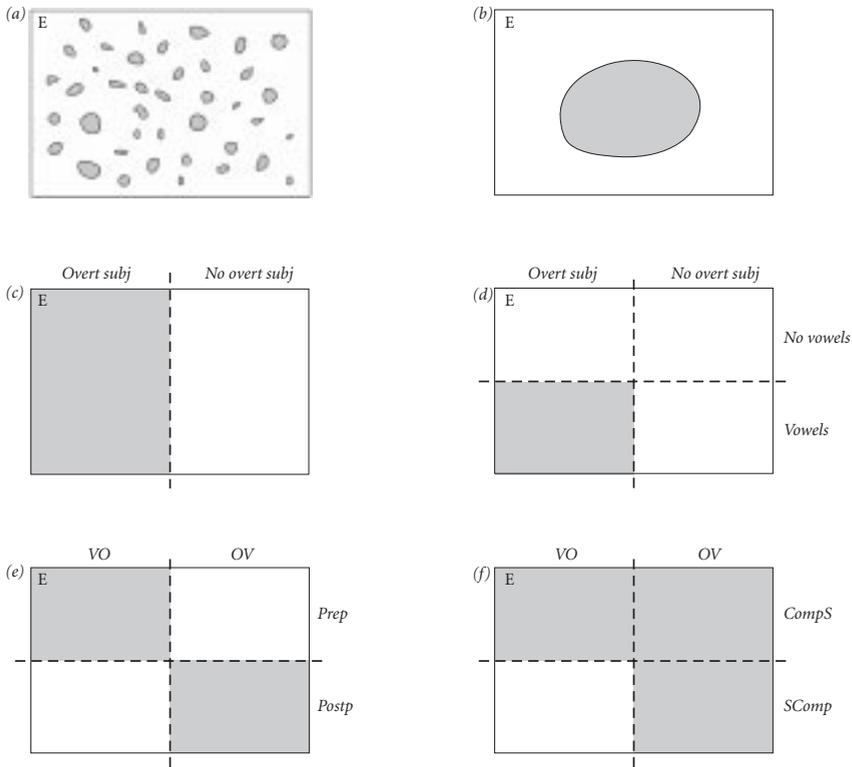


Figure 1.1. Venn diagrams showing various ways in which human languages (in grey) might vary in the space of logically possible languages (E). Refer to the text for details.

languages at the bottom of the space are those that do. Now, because there are no languages that lack vowels, the grey space in the figure is more tightly constrained. In theory typologists could go on adding more dimensions and get closer to a specification of ‘what makes a possible human language’.

However, there is something unsatisfying about this approach. Are we not just compiling a list of facts about language that are completely independent of each other? It is more interesting—and begs more explanation—when the orthogonal typological dimensions are actually related to one another in some way. Figure 1.1e gives just such an example. The left–right split in the space is between languages that typically order their verb before their object noun phrase (such as English in Example 1.1), and languages that order their object before their verb (such as Japanese in Example 1.2).

(1.1) Everyone loves someone

(1.2) Minna ga dareka o aisiteiru
all someone loving
'Everyone loves someone' (Shibatani 1990: 261)

Notice that, by giving two example languages, we can see that if this was our only typological dimension we would not be able to say anything at all about constraints on variation: both types occur.

The top–bottom split in the space divides languages with prepositions (preceding the noun phrase they are modifying, such as English in Example 1.3) from languages with postpositions (following the noun phrase they are modifying, such as Japanese in Example 1.4).

(1.3) Dr Heycock is teaching Japanese in the classroom

(1.4) Kakehi sensei ga kyoositu de eigo o osie-teiru
prof. classroom in English teach-be
'Prof. Kakehi is teaching English in the classroom'
(Shibatani 1990: 287)

Once again, there are examples of both types of language, so either typological dimension on its own tells us nothing about cross-linguistic distribution. However, if they are put together, as in the figure, then it becomes obvious that these two dimensions are related, in that the verb-before-object (or VO) languages are all prepositional and the verb-after-object (or OV) languages are all postpositional.²

There is another significant way in which two typological dimensions can be related. In Figure 1.1*f* the first dimension is the same as in the previous example—in other words, the split between VO and OV. The orthogonal dimension here, however, is the ordering of complementizer (e.g. English *that*) and subordinate clause. We can see that if a language orders its verb before its object then it will order its complementizer before the subordinate clause, as in the English Example 1.5. On the other hand, if a language is OV, then no prediction can reliably be made about the order of its complementizer and subordinate clause. So, whilst Japanese in Example 1.6 has a final complementizer, German (which is OV in subordinate clauses at least) has an initial complementizer in Example 1.7.

² There are some, though relatively few, exceptions to this generalization. According to Hawkins (1983), out of a sample of 336 languages, 26 are postpositional and VO, and 12 are prepositional and OV.

- (1.5) Ken knows that Heather sings the song
- (1.6) ken-wa heather-ga sono uta-wo utatta to itta
 Ken Heather that song sang COMP said
 ‘Ken said that Heather sang that song’
- (1.7) Ken weiss, dass Heather das Lied singt
 Ken knows COMP Heather the song sings
 ‘Ken knows that Heather sings the song’

These three languages exemplify points in the top left, bottom right, and top right of the diagram respectively. However, there are no languages that have VO&SComp, which is why the bottom left of the diagram is not filled in.

Forms of constraints

We have looked informally at the ways in which languages may vary in the space of logical possibility, and seen that there are many possible ways we can express the constraints on this variability. The formal expression of these cross-linguistic universals involves two distinct steps:

Typology. This is a taxonomy which categorizes 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.

The constraints may take various forms, which can be usefully categorized on two dimensions (notice that the broad distinctions here are recognized by other authors (e.g. Greenberg 1963; Comrie 1981; Hawkins 1988; Croft 1990), although the precise formulation is not identical). First, the constraints may be *absolute*³ 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 realize

³ The term *absolute universal* is sometimes used, by others, for substantive or formal universals that simply constrain languages so that they all have a certain property.

that a statistically significant skewing of the distribution of languages is as worthy of explanation as an absolute one.⁴

Secondly, the format of the constraint can typically be characterized 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:

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

where P_i is some property of a language L ⁵ that differentiates between a type T_i and T_i' , where a prime here indicates an opposite type.⁶

This logical statement can be paraphrased in prose as something like: ‘for all languages, if a language has property 1, then it will have property 2, and vice versa. Furthermore, if a language has property 2, then it will have property 3, and vice versa. In fact, a language either has all the properties from 1 to n , or none of them.’

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

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

Again, this logical statement can be paraphrased in prose as: ‘for all languages, if a language has property 1, then it will have property 2, but not necessarily vice versa. Furthermore, if that language has property 2, then it will have property 3. In fact, a language which has some numbered property will have all the properties from that number up to n .’

⁴ 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 categorizes languages into three 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 is 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.

⁵ For convenience we can simply abstract away from L in the expression of these universals in other places in this book.

⁶ This formulation relies on a binary typology. However, other typologies can be easily reduced to this case.

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 last two diagrams in Figure 1.1 (*e* and *f*) are examples of parametric and hierarchical universals respectively. The first universal parametrically relates verb/object order and adposition/noun-phrase order. This can be written as $VO \leftrightarrow Prep$, where $VO' \equiv OV$ and $Prep' \equiv Postp$. A prose paraphrase of this universal would be something like: 'All VO languages are Prep languages, and vice versa.' The second universal is different in that it rules out only one of the four logically possible types, $VO\&SComp$. This universal can be written, $VO \rightarrow CompS$, where $CompS' \equiv SComp$. This can be paraphrased: 'All VO languages are CompS languages.'

These two universals are examples of the simplest kinds of parametric and hierarchical universals, relating only two binary types. The first constrains languages to 2 out of 4 possible types, and the second to 3 out of 4 possible types. In general, parametric universals constrain attested languages to 2 out of 2^n possibilities, and hierarchical universals constrain to $n + 1$ out of 2^n . So, even for a small number of types, these universals are highly predictive.

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 (1963: 78) universal, 'languages with dominant VSO [verb–subject–object] order are always prepositional', can be expressed as $VSO \rightarrow Prep$. We could also rewrite this as a hierarchy $Prep > VSO$, and English could be placed halfway 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.

Table 1.1. *Grammatical number on nouns*

Plural	Dual	Trial/paucal
–	–	–
+	–	–
+	+	–
+	+	+

A well-known example of a hierarchy is given by Croft (1990: 96–8),⁷ referring to the possible expression of grammatical number on nouns:

plural > dual > trial/paucal

This corresponds to the universal:

(trial/paucal → dual)&(dual → plural)

In other words, if a language marks trial or paucal number on nouns, it will also mark dual; if it marks dual, it will mark plural. This hierarchy constrains human languages to four out of eight possibilities (adapted from Croft 1990: 97):

1. languages with only one noun form (languages without a category of number);
2. languages with singular and plural forms for nouns;⁸
3. languages with singular, dual, and plural forms for nouns;
4. languages with singular, dual, plural, and trial/paucal forms for nouns.

Another common way of visualizing this hierarchical universal is as a table where each row is a possible human language and + means that a language is of a particular type (see Table 1.1).

⁷ Note that Croft uses the < operator where we will be using >. Croft's use of this operator reflects relative *markedness*, a typological property to which we will return briefly in the next chapter.

⁸ Notice that, as it is presented here, the hierarchy makes no mention of singular number. It makes sense to talk of nouns having a number form only if there is something to contrast it with. So, it would be meaningless to say that a language had a plural form for nouns and no other form. The singular form is not mentioned in the hierarchy and can be assumed to be the default with which the others are contrasted. Alternatively we could add the 'singular' at the top of the hierarchy (as Croft does). Strictly speaking this would add an extra type of language to the predicted set: one that had no number at all (not even singular). However, this type cannot be distinguished from one that has only singular.

Table 1.2. *The contrapositive number hierarchy*

Plural'	Dual'	Trial/paucal'
–	–	–
–	–	+
–	+	+
+	+	+

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. English, for example, would be positioned on the second row.

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:

$$\text{trial'/paucal}' > \text{dual}' > \text{plural}'$$

where the prime symbol on these types simply refers to a language which *does not* mark that number category on any nouns. In other words, if a language does not have plural number, then it will not have dual number; if a language does not have dual number, it will not have trial or paucal number. In Chapter 3, the choice between a hierarchy and its contrapositive 'twin' will be shown to reflect on its explanation. The contrapositive table of possible languages (Table 1.2) is simply a mirror image of Table 1.1 (here, English would be on the third row):

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).

Perhaps the most extensive and critical review of the structure of functional explanation currently available is Newmeyer (1998), who has this to say about the variety of explanations and the problems of defining what *functionalism* actually means:

Those who share the functionalist orientation differ in their basic assumptions far more than do those who are committed to the generativist approach. This is partly

a consequence of there being a lot more possible ways that one can be against some theoretical conception (the compartmentalization of form) than one can be for it. Saying that the formal properties of language are not characterized as a system unto themselves leaves open a myriad of possibilities as to how they *should be* characterized. (§5)

I . . . characterize as 'functionalism' any approach that embodies the following three positions. . . . First, the links between formal properties of grammar and their semantic and pragmatic functions are tight enough to preclude any significant methodological or analytical 'parcelling out' of form. Second, to a significant degree, the formal properties of grammar are motivated by the functions that language carries out, in particular its function of conveying meaning in communication. And third, by means of integrating functional explanation with typological investigation, one can explain why certain grammatical features in the languages of the world are more common than others and why, for particular languages, the appearance of one feature often implies the appearance of another. (§5.4)

Newmeyer's definition of the functional approach is consistent with the characterization used in this book, where functionalism is seen as being concerned with explaining the structure of language (in particular language universals) by finding evidence of fit between that structure and language use.

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: 26–9; Hawkins 1988: 8–18; Hurford 1990: 94–6; Croft 1990: 252–6; Hall 1992: 27–32; and references therein). In this section we will look at some well-known examples that appeal to four rather different perspectives on use.

First, 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. Within the same English utterance, different instances of *I* or *me* always refer to the same entity. Similarly, almost all instances of *we* or *you* will refer to the same thing (unless the speaker points at different people during the utterance). 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 making co-referentiality unambiguous in third person referents than first or second person referents.

Another type of explanation appeals to *iconicity*, or the isomorphism of sign and signified. One of Greenberg's (1963: 93) universals 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. For example, in the English word *computations*, the derivational affix *-ation* comes before the inflectional affix *-s*. 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: 176) 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'.

A third type of explanation appeals to the structure of *discourse*. An interesting and complex example is DuBois's (1987) explanation of the tendency for languages' case systems to pattern as *nominative-accusative* or as *ergative*. 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 argues, using 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 noun phrases 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.

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* 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 cross-linguistic difference in the distribution of suffixes and prefixes 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 these features are all very different. 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: 67–8, 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 utilized 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 and 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 pronominalized is presumably related to the relative effort it takes for a hearer to recover the referent for a given versus 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 is probably the case for most. The rest of this book will therefore deal with explanations that appeal to pressures on production and perception of language, and I believe that the approach will be relevant to all functional explanations. Another reason to concentrate on this aspect of language use 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.

UG and universals

As mentioned earlier, the functional approach to explaining language universals contrasts 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.⁹ One of the purposes of this book, 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 them somewhat misleading. First, just as with the functionalist approach, these terms suggest an unwarranted degree of coherence. There are currently several broad theoretical programmes to which these labels could be applied. For example, *Principles and Parameters* (or *Government and Binding Theory*) (Chomsky 1981), the *Minimalist Program* (Marantz 1995), and *Optimality Theory* (e.g. Grimshaw 1997). 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 formalization 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 and Sag 1994) is the most popular at the moment. On the other hand, these theories could not really be called 'Chomskyan'.

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.¹⁰ If this is so, then we have a ready-made explanation for universal properties of languages (Hoekstra and Kooij 1988).

It seems then that the *innatist* and functionalist approaches are inevitably in competition as explanations of language universals. It is important to realize, however, that the central question that each approach is attempting to answer is different. Simplifying the situation drastically,

⁹ The syntactic (as opposed to phonological) bias of this book should be clear by this stage. The following review ignores the corresponding tension between functional and generative approaches to phonology.

¹⁰ This is not *necessarily* the case, of course. It is possible that some degree of variation in innate knowledge of language may be uncovered.

the difference can be characterized 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?’

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 do the constraints on variation have 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 are rich enough for language to be acquired using general-purpose learning mechanisms.’

The richly structured, innate UG or LAD 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).¹¹

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. First, a theory is *descriptively adequate* if it goes beyond a description of the linguistic data

¹¹ 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.

and accounts for a native speaker's intuitions about the grammaticality of utterances. In order that it can do this it must take into account that language has 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 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 be achieved only given innate syntactic knowledge. Clearly, not all 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

UG, 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)

We should treat with some caution claims that features of language are unlearnable without the particular type of innate knowledge embodied in the principles and parameters approach. However, it is clear that, at the very least, any kind of domain specific knowledge will aid the acquisition process.

Constraints on variation

Putting the learnability issue aside, what types of constraints on variation can this theory explain? First, the principles of grammar can directly constrain languages to be all of a certain type. For example, the universal that languages allow sentences to have subjects is trivially predicted from the extended projection principle, which includes a requirement that clauses have a position for a subject.

Secondly, 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 switching of settings in acquisition 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: 450–1) determines the overt-ness of *wh*-movement in a language. English has overt *wh*-movement (that is, expressions like *which*, *who*, and so on visibly move to the front of a clause), whereas Chinese has non-overt *wh*-movement (the equivalent expressions in this language appear *in situ*, but are assumed to move at some level in order that they can be interpreted correctly). 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 counter-intuitive 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 and Wexler (1987) propose a five-valued parameter which inputs into a definition of a governing category:

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 parameter.

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 and 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 likely to involve not 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 as opposed to absolute 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.

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.

First, 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 partially to alleviate the problem of learning language. As a side effect, Greenberg's (1963) universal mentioned earlier 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 generalizations about language she asks, 'how do such generalizations 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: 21–2) 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 biology.

However, the philosophical analogy between linguistic functional explanations and biological adaptation is not always fully worked out in linguistics.

To be completely explicit, we can formulate the following problem:

The 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 book 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 book are therefore formalized 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 the general answer to the problem above separate from the specific examples of the explanatory approach (e.g. the accessibility hierarchy and Hawkins's (1994*b*) performance theory). The



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

former is encoded as a simulation platform, and the latter as the particular initial conditions of a simulation run.

Overview

The rest of the book divides roughly into two parts. The first half goes into a theoretical approach to the problem of linkage and shows how this approach can be modelled computationally in order to test its validity with respect to particular explanations in the literature. The latter half of the book then reflects on the implications of the proposed approach for typology, functional explanation, and particularly innate theories of language variation.

The following chapter builds up a picture of the link between universals and function by considering in some detail Hawkins's (1994*b*) 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 characterizations 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's performance metric. It is shown that this model gives us a mechanism by which universals emerge, and as a bonus derives the prototypical time course of language 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 (AH). To do this, certain changes need to be made to the model to allow for multiple stable types to coexist. Once again, Hawkins's (1994*b*) performance theory is applied to the task, but the initial results are disappointing. It is argued instead that Hawkins's 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 relativization available to speakers. Finally 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 a case where parallel function apparently *is* expressed in a language. 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 book 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 book.