Introduction
After a careful definition of default inheritance, the paper proposes that this is the correct explanation for every kind of mismatch. In support of this suggestion there is a fairly detailed discussion of word order mismatches in English, followed by a brief re-analysis of the word-order data from Zapotec that Broadwell has used as evidence for Optimality Theory. The rest of the paper surveys 19 different kinds of mismatch in order to show how they can be treated by default inheritance as exceptions to defaults. The last section considers implications for the architecture of language.

1 Why DI Is The Default Explanation For Mismatches
My theoretical point in this chapter is basically a very simple one: if mismatches are departures from a default pattern, then the best mechanism for both stating and explaining them is default inheritance (DI). If this is true, there is no need to invoke any other kinds of mechanism for mismatches such as:

- Special procedures for converting the default pattern into the mismatch pattern by moving, deleting or otherwise changing it—the standard tools of derivational theories;

* I received very helpful comments on an earlier draft of this paper from Joe Hilferty, Adele Goldberg, Laura Michaelis, Elaine Francis and two anonymous reviewers. I should like to thank all of them.
- Special procedures for resolving conflicts between competing patterns such as the constraint ranking of Optimality Theory.

There are already several linguistic theories in which DI plays an explicit and important part:
- Cognitive Grammar (Langacker 1998; Langacker 2000)
- Some versions of Construction Grammar (Goldberg 1995; Goldberg this volume; Lakoff 1987)
- Some versions of Head-driven Phrase Structure Grammar (Flickinger 1987; Malouf 1998; Sag 2001)
- Word Grammar (Creider, Hudson 1999; Fraser, Hudson 1992; Hudson 2000a)

However there are others in which it has no place at all, so the claims of this paper are a major challenge to these theories. The role of DI raises fundamental theoretical issues which this chapter can hardly touch on, but at least I can explain how DI works and how it applies to some familiar examples of mismatch. This section will set the scene by explaining what DI is and why it is the best candidate for explaining all kinds of mismatch.

The basic idea of DI is extremely simple and familiar: Generalisations usually apply, but they may have exceptions. Any linguist is all too aware of this basic truth, but it is not restricted to language. The standard discussions of DI tend to start with examples such as three-legged cats (which violate the generalisation that cats have four legs), flightless birds and white elephants. In each case there is a generalisation about some super-category (e.g. Cat) which applies to the vast majority of its sub-cases: this is the ‘default’ pattern, so-called because sub-cases have it ‘by default’, i.e. unless something else prevents it. (Similarly, the default settings on a computer are those which apply unless someone changes them.) Thus if you know that something is a cat, but you cannot see its legs, then you can assume it has the default number: four. The logic which leads to this assumption is called ‘inheritance’—your particular cat inherits four legs (in your mind) from the general Cat. However, a small minority of cats have fewer than four legs, perhaps as the result of an accident, so there are exceptions whose actual characteristics ‘override’ the default. When your particular cat stands up and you can count the legs, you do not have to revise its classification—it is still a cat, but an exceptional one.

In short, the logic of DI allows you—as a human, and even as a scientist—to have the best of two worlds: a very rich and informative set of generalisations, but also faithfulness to the way the world actually is. Because your three-legged cat is a cat, you can still assume all the other default cat characteristics—purring, positive reactions to being stroked, and so on. This kind of inference is an absolutely essential life skill as it allows you to go beyond observable characteristics by guessing the unobservable ones. But it also allows you to be sensitive to the complex realities of experience by recognising and accepting exceptional characteristics. The inheritance system allows you to acquire vast amounts of information very fast—for example, a mere shadow is sufficient to warn you of a richly specified person or object—but the price you pay is constant uncertainty because the guessed defaults may, in fact, be overridden in the particular case (What you believe to be a cat that enjoys being stroked may turn out to hate it.).

It is important to stress that DI is merely a logic, and that like other logics it can produce false conclusions when applied to false premises. The classification may be wrong—the shadow may not belong to a human, the thing may not be a cat, classifying a dolphin as a fish is tempting but wrong. In the same way, the assumed default properties may have little or no basis in fact—one thinks of the assumed properties of ghosts. And, as just noted, the conclusion may be false because of an overriding property of which we were not aware. But these dangers are part of everyday experience and the virtue of DI is to be able to explain both the strengths and the weaknesses of everyday reasoning. There is strong evidence that we use DI outside language—that default reasoning leads us to ‘reconstruct’ memories of events (Barsalou 1992:141); that it plays an important part in vision (Luger and Stubblefield 1993:382); that we use it when deciding whether or not a speech event counts as an example of lying (Sweeters 1987); and so on.

It is easy to think of areas of language where the same logic applies. However it will be important to distinguish between typological mismatches and within-language ones. Typological mismatches are known only to typological linguists, who discover a general trend to which there are exceptions; e.g. the very strong tendency for subjects to precede objects in basic word order, to which there are a handful of exceptions. Presumably the linguists hold the facts in their minds as default patterns, but the facts are obviously independent of what linguists know about them. DI may or may not be a useful kind of logic in scientific work such as linguistic typology, and it may only or may not be right to postulate defaults as part of an innate Universal Grammar (Briscoe 2000); but this is a different issue from the status of DI within a grammar.

This paper will have nothing to say about typology so that we can concentrate on within-language mismatches. For example, the inflectional morphology of a single language typically shows general default
patterns which apply to most words, together with exceptional patterns which override these for certain irregular words. I shall try to show in this paper that DI extends well beyond morphology, and can be applied (I shall claim) to every known type of within-language mismatch. Indeed, I shall go further and claim not only that DI is suitable for every kind of mismatch, but that no other kind of logic should be used even if others can be made to fit the facts. I shall now try to justify this exclusive position.

Suppose some mismatch pattern in a language can be explained either in terms of DI or in some other way, such as by a set of basic patterns and a procedure for changing them. For example, we might imagine a procedure which defines the normal morphological pattern for plural nouns (stem + {z}), which when applied to goose gives geese, combined with a procedure for changing gooses into geese. The claim of this paper is that we should prefer the explanation in terms of DI unless there are strong empirical reasons for preferring the other (a situation which, I guess, never arises). Why should we always prefer the default-inheritance explanation? Here are some reasons:

- As explained above, DI is psychologically plausible, because we have good reasons for believing that it is part of our general cognition (How else could we cope with three-legged cats?). Clearly, an account which invokes machinery which is already available within cognition is better than one which invokes new kinds of machinery because it explains the phenomenon as an example of a more general pattern.
- As I hope to show below, DI fits all kinds of mismatch within language, rather than just one kind (e.g. inflectional morphology), so it again provides a more general explanation than alternatives which are restricted to a single area.
- DI is a logic which allows us to infer facts from a purely declarative database—a database consisting of ‘static’ facts such as “Cats have four legs” or “The plural form consists of the stem followed by [Z]”. This makes it ideally suited for a model of language in which the language itself consists of a declarative database of ‘facts’ which can be exploited equally efficiently by the processors responsible for hearing and for speaking. In this model, DI provides the logic which allows the processor to apply the general facts to more specific cases.
- As defined in the next section, DI is logically ‘clean’ in spite of widespread doubts (which we shall consider).
- DI is a formalisation of traditional ‘common-sense’ accounts of language patterns which go back over two thousand years to the Indian and Greek grammarians who managed to separate broad default generalisations from the exceptions that override them; so we need strong reasons for rejecting it in favour of a different logic.

All that I have tried to provide so far is an informal description and justification of DI. Later sections will show how it applies to mismatches in a number of different areas of language, but these explanations will need more specific and technical underpinnings which will be provided in the next section. From now on, I shall focus on how DI is handled in one particular theory: Word Grammar (WG).

2 The logic of DI

DI assumes a collection of categories which are arranged in a hierarchy called an ‘inheritance hierarchy’ which allows lower items to inherit characteristics from higher ones, and (if need be) to override these characteristics. Inheritance hierarchies have to show which of two related categories is the super-category and which is the sub-category (or member or instance). The literature contains a number of different ways of organising and displaying inheritance hierarchies, and the terminology varies from theory to theory; but I shall make the simplest possible assumption: that there is just one relationship which underlies all inheritance hierarchies. As in much of the Artificial Intelligence literature, I shall call this relationship ‘is-a’; for example, the word CAT is-a Noun, which is-a Word. I shall use my own notation for is-a links: a small triangle whose base rests on the super-category and whose apex is linked by lines to any sub-cases. Figure 1 shows a simple inheritance hierarchy using this notation. The two diagrams are exactly equivalent in terms of the information they display because the is-a links are not tied to the vertical dimension, but to the position of the triangle’s base.

---

1 A great deal of information about Word Grammar is available at (or via) http://www-pixen.net.ac.uk/home/dick/wg.htm. The main monograph-length treatment is Hudson 1990.

2 Like other users of the term ‘see’ I use it as a verb, but I add the hyphen (is-a) for convenience in order to allow the first part to inflect in the usual way: is-a, are-a, isn’t-a, be-a, and so on.

3 I use initial capitals for the names of stored concepts. Thus the name of the concept ‘cat’ is Cat, whereas a member of this concept is simply a cat. Similarly, noun test Word but a noun is a word. For lexemes I use uppercases throughout, so CAT is-a Noun, and the sense of CAT is Cat.
An inheritance hierarchy must be part of a larger network of information which also contains the facts that are available for inheritance—the 'defaults'. For example, suppose (for the sake of a simple example) we wish to show that a word's stem is normally a single morpheme. Information such as this can be represented in the form of a network built around the inheritance hierarchy, in which labelled nodes are connected to one another by labelled links. Figure 2 shows how the necessary links can be added to the inheritance hierarchy of Figure 1. The unlabelled dot stands for a variable—i.e. something whose identity varies from occasion to occasion; so in words, the typical Word has a stem which is-a Morpeme.

This little network illustrates an important characteristic of WG networks, which is that all the links are classified. Even here we have two distinct link-types: is-a and Stem (short for 'stem-of'), and we shall meet a large number of other link-types. Indeed, we shall see below that link-types are themselves organised in an is-a hierarchy just like the hierarchies that organise the entities that they connect—for example, Stem is-a Part. WG makes use of a great many familiar relationships—Sense, Dependent, Subject, Speaker and many more—and can easily accommodate new relationships. This is not a weakness of the theory; it reflects a claim of the theory which is that (most) link-types are like (most) entity concepts in being learned rather than innate. If there are universal limits on possible relationships, they need to be discovered by research rather than imposed by the theory.

We can now apply DI to the little network in Figure 2 in order to let the information about stems spread down the hierarchy, first to Noun and then to CAT. This is a very simple and mechanical copying operation in which information is copied from Word to lower nodes, but there is an important detail to be noted. The lower node cannot simply link to the same variable as Word, because this would mean that different lexemes (e.g. CAT and DOG) would both link to the same variable and therefore must have the same stem, which is obviously wrong. Instead, we create a new variable node which is-a the higher one; and for reasons which are similar but less obvious, we do the same for the links so the 'stem' link from Noun is-a that from Word. The result is the arrangement in Figure 3, where inherited links are shown as bold arcs. In words, because a noun is-a word and a word has a stem, the noun also has a stem, and similarly CAT has a stem because Noun has one. In each case, the link is-a the one which is inherited, and the same is true of the variable node. Consequently, each of the stems is-a morpheme. These is-a links allow other words to inherit similar characteristics without all (wrongly) inheriting the same morpheme.

This very elementary example provides the formal basis on which we can start to build an efficient logic for DI. In particular, we can treat the classified arcs as defining functions in the mathematical sense of a relationship from the argument to a unique value; so Stem is a function

---

4 Some people object to the idea that a typical word has a monomorphic stem, i.e. the claim that HAPPY is a more typical word than HAPPINESS. The evidence is that learners must, surely, assume that new words are monomorphemes unless they have evidence to the contrary (in terms of partial form-meaning correspondences to existing words). For example, when we first heard ELEPHANT we had no reason to divide it into morphemes so we stored it as a monomorpheme. We often learn to impose morphological structure on words which we have considered monomorphemes, but the revision rarely goes in the opposite direction. This is as expected if the default assumption, made in the absence of evidence to the contrary, is that a word's stem is a single morpheme.

5 In a network analysis, identity is shown entirely by the geometry of the whole network rather than by labels (Lamb 1966; Lamb 1998); so if we want to show that two concepts are in fact the same, we cannot do this by attaching the same label to two distinct nodes. Instead, we have to merge them into a single node, and we show the similarities between them by an is-a link. By the same logic, a single relationship can only connect one pair of nodes; so two separate arcs in a network diagram must in fact represent distinct relationships whose similarities are again shown by an is-a link.
from a word to its (unique) stem morpheme. In this formulation the expression “the F of A is B” means that there is a function F whose argument is A and whose value is B—in other words, there is an arrow called F pointing from a node A to another node B.

[1] Default Inheritance Axiom (Preliminary version)
If:  \( A \rightarrow a \ B \), and the F of B is C,  
then: the F' of A is C', where: F' is-a F, and C' is-a C.

The axiom is presented as a network in Figure 4. This configuration is not tied to any particular content, so it matches any part of any network and allows inheritance freely.

The example just given assumed that the stem of CAT has to be inherited from Word, but this is unrealistic because we all know that it is not just some morpheme, but specifically the morpheme \{cat\} (The example would be realistic in the situation where we cannot recall the stem of a particular lexeme—a common situation for most of us. At least we know that what we are looking for is a morpheme, rather than, say, a complete sentence or a fully inflected word form). We can add this information to give Figure 5 without introducing any conflict with the information already provided in Figure 3. Notice that the stem of CAT is an example of \{cat\}, rather than \{cat\} itself, because there are other uses of \{cat\} such as in the stem of the adjective CATTY.

The crucial part of DI for the treatment of mismatches is the overriding of defaults. Take the word CATTY for example; what is its stem? In a fuller analysis of morphology, the stem is the form to which we add inflectional suffixes such as the comparative suffix \{er\}, as in catter (meaning ‘more catty’), so the stem of CATTY must be \{catty\}; but we have just seen that \{cat\} is only one part of this form, so it must be an exception to the default rule in which the stem is not a morpheme, but a complex form containing two morphemes: \{cat\} and \{y\}. Figure 6 includes this irregularity (and loses some of the detail of Figure 5). In prose, a word has a stem which is a morpheme, and CAT conforms to this generalisation by having (an example of) the morpheme \{cat\} as its stem; but CATTY exceptionally has a complex form which consists of (examples of) the two morphemes \{cat\} and \{y\}. 

![Figure 3](image1.png)

![Figure 4](image2.png)

![Figure 5](image3.png)

![Figure 6](image4.png)
As it stands, the D1 axiom does not make provision for overriding, so it needs to be revised:

[2] Default Inheritance Axiom (Final version)

If: A is-a B, and the F of B is C,
then: the F' of A is C',
where: F' is-a F, and C' is-a C,
unless: the F' of B' is D,
where: A is-a B' or A = B',
B' is-a B,
D isn't-a C.

In short, A inherits B as the value for F' unless it can get a conflicting value ("isn't-a C") from a concept B' lower down the hierarchy, including A itself. The logic gives an unambiguous definition of the situations in which one fact overrides another:

- Both facts have the same logical structure Function (Argument) = Value.
- The overriding's Function must is-a that of the default.
- The overriding's Argument must is-a that of the default.
- The overriding's Value must not is-a that of the default.

All the relationships are extremely local as they involve small segments of just three inheritance hierarchies:

- The is-a links between the overriding Function and the default Function.
- The is-a links between the overriding Argument and the default Argument.
- The answer to the simple query about whether the overriding Value is-a the default Value.

It is always easy to find a winner among competing inheritable values by checking these limited is-a relationships. For example, in Figure 6, the complex form \{catty\} wins over the default 'some morpheme' because they are both possible values for the function 'stem'; and CATTY is below Word in the inheritance hierarchy.

The revised axiom is presented as a network in Figure 7, where an irregular item A1 is contrasted with a regular one A2.

What this diagram does, in effect, is to allow an item to be either the same as its super-category or different— in other words, virtually anything is possible in principle. This may seem much too liberal:

- If we define a penguin as a bird that does not fly, what is to prevent us from asserting that a block of wood is a bird that does not fly, does not have feathers and does not lay eggs? (Luger and Stubblefield 1993:389)

The answer to this question is intuitively obvious: there is no point in asserting this because the block of wood will inherit nothing from Bird. In short, we only classify something as a bird if it has more features in common with birds than with anything else—even a penguin.
is more like a bird than anything else if we consider its characteristics globally. This basic principle of classification has been called the Best Fit Principle (Hudson 1984; Winograd 1976:20) and is clearly fundamental; but for lack of space, we shall have to take it for granted here. In brief, it can be thought of as global pattern-matching—matching all the known facts of the current bit of experience, including context, onto the knowledge base in such a way as to minimize discrepancies.

DI is contentious (Dabrowska 2001; Green, Morgan 1996; Kay, Fillmore 1999; Shieber 1986) because the logic is ‘non-monotonic’ or ‘defeasible’—conclusions based on a subset of premises may be overturned by later premises, so the growth of reliable inferences is not simply additive (‘monotonic’). Worse still, every inference may (in principle) be overridden, so the system cannot rest until it has hunted exhaustively for overriding facts—where ‘exhaustively’ may mean the entire database of stored facts, or even an infinite set of facts which could be inferred from these. If this procedure really is essential in DI, then it surely constitutes proof against DI. Alternative conclusions are available in the literature of logic, however (Antoniou, Williams 1997; Bernard 1989; Cholewinski et al 1999; Faron, Ganascia 1997; Linke, Schaub 2000), and the objection clearly does not apply to the logic described above. There is no need to search the entire database, let alone all its consequences, because overriding facts are always local. A fact about function $F$ of $A$ which is inherited from function $F'$ of $A'$ must involve a function between $F$ and $F'$ and a concept between $A$ and $A'$ in their respective inheritance hierarchies—a very limited search space indeed. This definition of DI is similar to the definition used in the inheritance language DATR (Evans, Gazdar 1996), which has been exploited successfully in various computer simulations of natural language including Network Morphology (Corbett, Fraser 1993; Fraser, Corbett 1997; Hippisley 1998). Finally we must recognize the potential difficulties of multiple inheritance. Inheritance from multiple super-categories is already allowed by the DI axiom, but the problem lies in the potential conflicts which it can produce. One well-known example is known as the ‘Nixon diamond’: the American President Richard Nixon was both a Quaker and a Republican; as a Quaker he rejected war, but as a Republican he accepted it. These properties are both inheritable but they conflict (Touretzky 1986). How should we react? The only reasonable reply is that such conflicts are a fact of cognitive life and we, as cognitive scientists or linguists, must learn to live with them. Nixon lived with the conflict, but resolved it by giving priority to one of the alternatives—in other words, he assigned himself the Republican value by fiat. His mental state is diagrammed in Figure 8: Nixon is a Quaker and also is a Republican; Quaker and Republican both is a Person; Person has an undefined value for ‘war’—a place-holder for a more complex analysis of attitude to war. For Quaker the value is No, but for Republican it is Yes, and Nixon should logically inherit both but stipulates Yes, which overrides No.

As far as language is concerned, multiple inheritance appears to be all-pervasive and cannot be avoided. An obvious area of application is in inflectional morphology, where words inherit both from a lexeme and from some inflectional category (or categories). For example, the plural of CAT inherits both from CAT and from the inflectional word-type Plural; from the former it inherits (inter alia) its stem, and from the latter its suffix. These are combined to form the word’s ‘whole’ (its fully inflected form), which by default is the same as the stem (Creider, Hudson 1999; Hudson 2000a). In contrast, the plural of PERSON is irregular (people) so it does not inherit from either super-category. The relevant part of the grammar is shown in Figure 9 (which is simplified by the omission of is-a links among relations). The regular plural pattern is shown around ‘Plural’, whose ‘whole’ has two parts which respectively are-a whatever the stem happens to be and the morpheme {Z}. The pattern for CAT:plural (the plural of CAT) is simply a copy of this, in contrast with that for PERSON:plural, which overrides it by defining a suppletive whole.
Infectional morphology is probably the most obvious area of mismatch between regular and exceptional patterns in language; after all, what we mean by irregular verb or irregular noun is a word which is irregular specifically in its infectional morphology (rather than, say, in its syntactic complementation). However the remaining sections will show that DI applies to other areas of language. Sections 3 and 4 will focus on word order in some detail, and Section 5 will consider a wide range of other kinds of patterns.

3 Mismatches between default and exceptional word orders: English

In syntax it is widely accepted that some word orders are more basic than others—that this assumption explains the facts better than an alternative in which all patterns have the same status. In derivational theories, these differences are explained in terms of movement rules; but DI is an obvious alternative. In a DI account, the basic orders are defaults which the more specialised patterns override. Word order is a good area for illustrating the formal properties of DI because there are clear cases where an exceptional order is itself overridden. Thus the word order example shows that priority depends on nothing but relative position in the inheritance hierarchy, rather than on some kind of stipulated classification as 'default' or 'non-default'. We shall also be able to show how important it is to be able to classify relations hierarchically.

The following discussion will assume an analysis of syntax in terms of dependencies between words rather than in terms of phrase structure. For example, sentence (1) will have the structure in Figure 10 rather than the more familiar one in Figure 11. In Figure 10, the arrows show dependencies, with the arrow pointing towards the dependent; each dependency shows some familiar grammatical relationship such as 'subject' or 'adjunct'. The contrast between pre- and post-head adjuncts will be explained below.

(1) The bald linguist wrote an article during the summer.

Arguably, dependency analysis has important advantages over phrase structure in all domains of analysis—theory, description, psycholinguistics, computation—but this is not the place to present them.6 However,

---

one advantage is highly relevant: dependency analysis makes DI very easy to apply because all generalisations apply to the inheritance hierarchy of words. In contrast, other theories (notably Head-driven Phrase Structure Grammar) have to reconcile an inheritance hierarchy of words with a hierarchy of phrase types (Pollard, Sag 1994; Sag 1997:31).

The bald linguist wrote an article during the summer.

**Figure 10**

We start with word order in English. It is widely agreed that English is a basically a head-initial language, as head-initial patterns dominate both in the grammar and in usage. The default pattern, therefore, is the one shown in Figure 12, and it is reasonable to claim that this pattern 'explains' the head-initial tendency of English syntax. In a network analysis, the left-right dimension has no significance, so linear order has to be treated like any other relationship. In WG notation, the relevant relationship is 'predecessor', labelled '<<'. The arrow points from the later to the earlier. To emphasise the irrelevance of the horizontal dimension, in this figure the dependent is shown to the left of the headword. In prose, then, Figure 12 says that a word typically depends on another word which stands before it (This is, of course, a parochial rule for English; there is no suggestion that every language has the same basic word order.).

**Figure 12**

In spite of the clear preference for head-initial patterns, English includes a significant minority of head-final patterns which are explained by a more specific generalisation.\(^7\) Given the word-class of the dependent and head and the grammatical relation concerned, the direction is generally fixed, so each dependency has a default direction. Moreover, the default direction determines the way in which various special patterns apply. For example, extraction applies only to dependents whose default position is post-head (with special arrangements for subjects of subordinate verbs) (Hudson 1996:354-403). All these facts point to a fundamental distinction between two types of dependent: pre-head dependent and post-head dependent, abbreviated to 'pre-dependent' and 'post-dependent'. The main characteristic of pre-dependents is to override the default head-initial pattern, as shown in Figure 13, where a word is by default the predecessor of its dependent, but exceptionally a word's pre-dependent is also its predecessor.

This distinction cuts across the other main distinction between de-

\(^7\) One reviewer argues that the simplest generalisation is in fact that lexical heads precede phrases. This is not clearly true. For one thing it presupposes a phrase-structure analysis in which VP is the head of a clause, which is clearly incompatible not only with a dependency analysis but also with Minimalist analyses (where most 'heads' are functional categories), and indeed with X-bar theory in general. For another it ignores adjective phrases used attributively before a head noun (e.g. *very big house*).
pendents that are selected by the head word (called ‘valents’) and those that are not (adjuncts). In short, the inheritance hierarchy for dependents allows multiple inheritance, whereby subjects (for example) inherit both from ‘pre-dependent’ and from ‘valent’. Figure 14 shows the cross-classification of dependents.

Figures 13 and 14 explain the mismatches in our example sentence between actual word order and the default head-first pattern. As can be seen in Figure 10, the two exceptions are the subject dependency between the (bald linguist) and wrote, and the pre-adjunct dependency (a left-pointing arrow) between bald and linguist. In both cases, the dependent is a pre-dependent, so it inherits the special pre-head word order of Figure 13 which overrides the default order.

Both of these exceptional word-orders can in turn be overridden by yet more specific rules. On the one hand, the subject may follow its parent if this is an ‘inverting’ verb—a tense auxiliary verb with the characteristics (e.g. interrogative meaning) that require subject-aided inversion. For example, the subject follows its parent in (2).

(2) Is he ready?

Such cases need a more specific rule which reverses the effects of the rule for pre-dependents, as shown in Figure 15. The logic of D1 guarantees that the pattern for subjects of inverting verbs overrides that for pre-dependents in just the same way that the latter overrides the one for dependents.

It is worth stressing that the inverted subject is still a ‘pre-dependent’ in spite of its position; ‘pre-dependent’ means only that the dependent takes a pre-head position by default—but, as usual, it is compatible with exceptions. This classification provides a non-transformational
equivalent of an ‘underlying’ word order which explains, for example, why extraction out of a subject is bad (barring parasitic gaps) whether or not it is inverted (Hudson 1990:191):

(3)  
    a. I don't know who he would accept a picture of.  
    b. *I don't know who a picture of would be acceptable.  
    c. *Who would a picture of be acceptable?

This is easily explained if extraction is restricted to post-dependents; in spite of its ‘surface’ position, the subject in example (c) is not a post-dependent, so extraction is questionable.

As mentioned above, similar word order mismatches are possible for attributive adjectives. An adjective which depends on a common noun is its pre-dependent, so by default the adjective comes first; but this is merely the default. A well-known fact is that if the adjective itself has a post-dependent, the adjective is put after the head noun: for example, alongside keen students we find students keen on syntax, not *keen on syntax students. (In contrast, a pre-dependent has no effect on the adjective’s position: extremely keen students.) Once again the change in word order does not alter the fact that the adjective is pre-dependent, which explains why it cannot be extrapolated:

(4)  
    a. Any student who likes syntax is welcome.  
    b. Any student with an interest in syntax is welcome.  
    c. Any extremely keen student is welcome.  
    d. Any student keen on syntax is welcome.

Example (c) shows that pre-dependents cannot be extrapolated, and (d) shows that this restriction extends even to pre-dependents which would follow the head noun.

In this case the special word order is not tied to a word-class (such as ‘inverting verb’), but to a type of dependent: ‘post-modified pre-dependent’, characterised both by the special word order and by the presence of a post-dependent. Thus a dependent follows the head word, a pre-dependent precedes it, and a post-modified pre-dependent follows it. The relevant part of the grammar is shown in Figure 16.

Finally we should look at ‘wh-movement’, as it is perhaps the best-researched example of mismatch in English word order and illustrates nicely the general tendency for deviant cases to have more complex structures than defaults. Like the earlier examples, the treatment builds on the distinction between pre- and post-dependents. We start with a simple example which involves nothing but special word order.

(5)  
    a. *Who am I?

    Both of the verb’s dependents are in non-default positions. I is a pre-dependent (subject), but follows the head-word because this is an inverting verb; and who is a post-dependent (complement) but precedes the head-word because it is an interrogative pronoun and needs to be initial. We already have an explanation for the position of I (see Figure 15), but not for that of who.

    We might consider an analysis for extraction similar to the one given above for subject inversion, in which we left the dependency structure untouched but changed the word order, making I a pre-dependent in an exceptional position. We supported this by showing that an inverted subject does not have the characteristics—such as free extractability—that other post-dependents have, so it is not a post-dependent. This kind of evidence is harder to find for {wh}-movement, and there are in fact reasons for thinking that an initial wh-item such as who is in fact a pre-dependent as well as a post-dependent. The main reason for thinking this is that the wh-item can be moved out of its clause by ‘long-distance extraction’ which can relocate it at any distance to the left.
(6) Who do you think other people say I am?

In an example like (6), it is clear that the position of who is not simply a reversal of its normal dependency-based position which would place it before am rather than after it; rather, who takes its position from the very first verb do, so there must be a dependency structure which justifies this, in which who depends on do as a pre-dependent. In short, the ‘landing-site’ of the extraction is defined as ‘pre-dependent of do’, while its ‘launch-site’ is ‘post-dependent of am’. For familiar reasons, the best way to connect the landing-site to the launch-site is via a succession of pre-dependency links to think, say and am; therefore the structure for Who am I? must be as shown in Figure 17, where who has two different dependency relations to am: complement (post-dependent) and a new dependency type, ‘extractee’, which is-a pre-dependent. The abbreviated label for ‘extractee’ is ‘x<’, where ‘<’ again stands for a left-pointing arrow.

The syntactic structure for Who am I? is shown in relation to the parts of the grammar that determine the order of the words, and illustrates the role of DI in word order. To recapitulate, both who and I are in non-default positions, but the grammar provides a different explanation in each case:

- For I, the default position is before the verb because I is its subject, which is-a pre-dependent, and the default order rule for pre-dependents overrides the default for dependents; however the verb is an inverting verb, so this exception is itself overridden.
- For who, the default position is after the verb because who is its complement, which is-a post-dependent so the default word order for all dependents should apply. However because who is an interrogative pronoun (a kind of wh-item), it is also the extractee of am, which means it is a pre-dependent. Since pre-dependent is-a dependent, the post-head position of pre-dependent beats the post-head position for dependent.

There is a great deal more that could be said about word order in English, but my general claim is that it can all be said in terms of a network grammar along the lines illustrated, combined with DI.

4 Mismatches between default and exceptional word orders: Zapotec

It is important to show that DI can cope with data that have specifically been claimed to require some other apparatus. For example, Broadwell has claimed that certain word-order facts from the Mexican language Zapotec, show the need for an analysis in terms of Optimality Theory (Broadwell 1999). The language is head-initial, with basic V S O clause order:

(7) U-diny Juany be’cw cun yag.
    com-hit John dog with stick
    'John hit the dog with a stick.'

The structure for this sentence is presumably as shown in Figure 18, which also shows the relevant word-order generalisation. (For convenience the Zapotec words are replaced in the following diagrams by their English equivalents; henceforward I ignore the Zapotec prefix i- in the English translations.)

As in English, interrogative pronouns are front-shifted, so we can assume that, as in English, they receive an extra dependency called ‘extractee’ and labelled ‘x<’ which requires exceptional pre-head order. An example is (8), whose structure can be assumed to be as in Figure 19.

Broadwell’s data relate to the dialect spoken in San Dionicio Ocotepec, in Oaxaca.
The general rule for wh-pronouns is therefore, as in English, that they depend on the verb as extracted in addition to any other dependency they may have (e.g. as object); the word order rule for extractee overrides the default for all dependencies, so wh-pronouns precede the verb.

So far, Zapotec fits easily into the patterns familiar from other languages. What makes it especially interesting is the treatment of more complex front-shifted phrases illustrated by the following examples:

(9) a. Xhį cını u-dίny Juany bč'cw?
   what com-hit John dog
   'What did John hit the dog with?'

   b. Tū x-pè'cw cū'ā Juany?
   who poss-dog com-grab John
   'Whose dog did John grab?'

   c. Tū bč'cw cū'ā Juany?
   which dog com-grab John
   'Which dog did John grab?'

In each case the order given is the only one possible within the fronted phrase, but in each case this is the reverse of the normal phrasal order, where prepositions usually precede their complement and possessors and demonstratives follow the head noun. Why might Zapotec allow these reversals of the default order, and how might they be explained formally?

It is easy to see intuitively why sentences like this are permitted as a way of keeping the wh-word close to the word on which it depends even when it occurs at the start of the clause. English has a similar pattern (‘prepositional pied-piping’) in which, unlike Zapotec, the normal order of preposition and wh-pronoun is preserved:

(10) For whom is this?

In both languages the peculiarity of the structure lies in the fact that the preposition takes its position from its complement (the wh-pronoun), contrary to the general principle that dependents take their position from the head, not vice versa. The solution for English lies in an additional dependency called ‘proxy’ (abbreviated ‘y’) whereby the preposition acts as a proxy for the wh-pronoun—i.e. as though it were itself a wh-pronoun (Hudson 2000b; Rosta 1997). This extra dependency is invoked by the rule for extractees: a verb’s extractee may be either a wh-pronoun or the proxy of a wh-pronoun.\(^3\) Figure 20 shows the structure for (10), with for as proxy for whom.

Zapotec seems to use the same facility for fronting prepositions along with their dependent wh-pronouns, but it also extends the facility to nouns with a dependent wh-word (whose dog, which dog). Let us assume, therefore, that in Zapotec any wh-word assigns the proxy role to any non-verb on which it depends. As in English, a wh-word’s proxy is the verb’s extractee. However, another peculiarity of Zapotec is the rule

\(^3\)A slightly more elegant formulation allows a word to be its own proxy, in which case extraction always applies to the proxy of a wh-pronoun.
For whom is this?

Figure 20

which ensures that the wh-word is clause-initial by reversing the normal order of head and dependent, giving *whose dog* instead of the expected *dog whose* and *what with for with what*. Thus the proxy relation has two effects on word order:

- It requires the wh-word's proxy to precede the verb (by virtue of the extractee relation);
- It requires the wh-word to precede its proxy.

As we shall see below, these two effects are separable; so we shall attribute them to different relations. The first is due to the proxy relation itself, while the second is due to a new relation called 'follower', and abbreviated as ‘>’ (suggesting a right-pointing dependency arrow). In most cases a word's proxy is also its follower, so in example (a), *with* is both proxy and follower of *what* and similarly for the other examples. The structure for this example is shown in Figure 21, together with the parts of the grammar which sanction the proxy and follower relations.

Figure 21

So far, then, we have seen that Zapotec is like English except that pied-piping also affects the order of local dependents in such a way that the wh-word is clause-initial. However, further complications arise when the interrogative word combines the two possibilities just discussed: it depends on a noun which in turn depends on a preposition (e.g. 'with which stick' or 'with whose stick'). In this case there are two possible orders, but all other logical possibilities are ungrammatical:

(11) a. Xhi cui yag ñ-diny Juany bë'cw?
   which with stick hit John dog
   'With which stick did John hit the dog?'

b. Cui xhi yag ñ-diny Juany bë'cw?
   with which stick hit John dog
   'With which stick did John hit the dog?'

These patterns can be seen as two attempts to fill the communicative gap which the grammar would otherwise leave; both attempts make some sense as a response to this functional pressure.

The order *which with stick* is a simple generalisation of the two-word pattern, which has the advantage of starting with the interrogative word and of preserving the normal order of *with stick*. The price that it pays is to separate *which* from *stick*, producing a discontinuous phrase. Figure 22 shows how the grammar generates this pattern by allowing the preposition to be the proxy of its complement's dependent (One arrow is drawn beneath the words simply for convenience.). As usual,
the wh-word’s proxy (the preposition) is also its follower, but the noun has to be positioned as well, so in this case the noun is also treated as the wh-word’s follower. Notice, however, that only one word can be the wh-word’s proxy because (for Zapotec, though not universally) only one word can be extractee, so here the noun is the wh-word’s follower without being its proxy.

![Diagram of wh-word movement](image)

**Figure 22**

The second pattern, with which stick, strikes a different compromise. It preserves the usual head-initial position of the preposition with and also keeps which before stick, but loses the clause-initial position of which. The difference is the result of just one difference compared with the alternative order: with is no longer the follower of which. The structure is in Figure 23. The grammar is the same as in Figure 22 except that the follower link from the wh-word to its proxy (the preposition) is optional.

![Diagram of wh-word movement](image)

**Figure 23**

We have now considered DI analyses for all the problem data from Zapotec which Broadwell uses as evidence for Optimality Theory. The analysis proposed here required three special dependencies for wh-movement:

- extractee, the basic dependency for front-shifted elements also found in English; this is responsible for putting an ordinary dependent in front of the verb;
- proxy, found (among other places) in prepositional pied-piping in English; this allows the words on which the wh-word depends to front-shift with it;
- follower, which may be unique to Zapotec, and which allows the wh-word to occur before words which it would otherwise follow.

DI plays an important part in the analysis by allowing extractee and follower to override the normal head-initial order.

5 Other mismatches in brief

The discussion so far has focussed almost entirely on mismatches of inflectional morphology and word order, but the same principles of default inheritance apply to any kind of mental generalisation, whether inside language or outside it, so we can expect them also to apply to the many other kinds of linguistic patterning that fall under the general description ‘mismatch’. In this section I shall present a rapid review of a number of examples which are deliberately chosen for their diversity. The available space will allow no more than a thumb-nail description.
of the mismatch and of how DI can explain it, with a picture of the relevant part of the grammar. The examples are summarised in Table 1.

**TABLE 1**

<table>
<thead>
<tr>
<th>LEVEL</th>
<th>EXCEPTION</th>
<th>DEFAULT</th>
<th>EXAMPLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>morphology, syntax</td>
<td>irregular morphology</td>
<td>plural = stem + [2]</td>
<td>people</td>
</tr>
<tr>
<td>morphology, syntax</td>
<td>clitics</td>
<td>A word-part is a morpheme.</td>
<td>you’re</td>
</tr>
<tr>
<td>syntax</td>
<td>merged words</td>
<td>A morpheme is part of one word.</td>
<td>me + a = my</td>
</tr>
<tr>
<td>syntax</td>
<td>raising</td>
<td>A word has one parent.</td>
<td>seen</td>
</tr>
<tr>
<td>syntax</td>
<td>gerunds</td>
<td>A word is a one-word-class.</td>
<td>doing syntax</td>
</tr>
<tr>
<td>semantics</td>
<td>exceptional valency</td>
<td>An auxiliary’s complement is a bare infinitive.</td>
<td>ought</td>
</tr>
<tr>
<td>semantics</td>
<td>passives</td>
<td>The ‘er’ is expressed by the subject.</td>
<td>was appointed</td>
</tr>
<tr>
<td>semantics</td>
<td>‘re-classifiers’</td>
<td>Modification only adds detail.</td>
<td>fake diamonds</td>
</tr>
<tr>
<td>semantics</td>
<td>idioms</td>
<td>Every noun belongs to one word.</td>
<td>hot dog</td>
</tr>
<tr>
<td>semantics</td>
<td>pluralia tantum</td>
<td>A plural denotes a set.</td>
<td>scales</td>
</tr>
<tr>
<td>semantics</td>
<td>prototype effects</td>
<td>An instance inherits all characteristics.</td>
<td>That cat has three legs.</td>
</tr>
<tr>
<td>semantics</td>
<td>mistakes</td>
<td>A word is as in the grammar.</td>
<td>mis spellings</td>
</tr>
</tbody>
</table>

**Irregular morphology.** Irregular morphology provided our first example of default overriding, with *people* as the irregular plural of *PERSON*. Figure 9 presented this exception in the context of the default morphology for plural nouns. It is relatively obvious how DI applies to most areas of morphology, as it underlies traditional and informal descriptions in terms of generalisations and exceptions. However DI is surprisingly absent from more formal treatments, and has been completely ignored as a way to formulate rules in the recent debate about the psycholinguistics of irregular inflection (Clahsen et al 1992; Pinker 1998; Pinker, Prince 1994; Prasada, Pinker 1993).

**Clitics.** A clitic is a word which is syntactically related to other words in the normal way but also morphologically related to one word as one of its parts. For example, in the sentence You’re wrong, what we write as ‘re is a verb, just like are in You are wrong, but it is also part of a larger word you’re. The words ‘re and you’re are both exceptions to the default generalisation that the parts of a word are morphemes, and they both override the default because they are below Word in the inheritance hierarchy (Hudson 2001). According to Figure 24, a word’s parts are typically morphemes, but one of the parts of the ‘hostword’ YOU’RE is the clitic ‘re, a kind of word.

![Figure 24](image)

**Merged words.** This term is meant to cover cases where two syntactic words are realised as a single word-form. (Word-forms are simply forms which may be shared by more than one word; for example the word-form run may belong to the plural of the noun RUNners or to the singular of the verb RUN.) For example, in French the combination de le, ‘of the’, where le is masculine singular, is replaced by the single word-form du (e.g. de la viande, ‘of the meat’, contrasts with du pain, ‘of the bread’). Similarly in English one can argue that my is the merged form of me’s, so that my hat actually has exactly the same dependency structure as John’s hat (Rosta 1997). The default pattern shows one word-form per word, but as shown in Figure 25, when a personal pronoun is combined with the possessive marker ‘s, the two share a single form; e.g. ME + ’s has a single word-form, my.

**Raising.** By default a word depends on just one word. However, the well-known ‘subject-raising’ pattern is handled in WG, as in other constraint-based theories, as an example of ‘structure-sharing’ (Pollard, Sag 1994:2). This means that the raised subject is related in the syntax to both the lower ‘source’ verb and the higher ‘raising’ verb; so in He seems to like her, he is the subject of like (and of to) as well as of seems. There is some evidence (Hudson 1984:117) for preferring a purely syntactic analysis rather than one which treats raising merely as a mismatch between syntax and semantics (e.g. Sadock 1991). As shown in Figure 26, the raising pattern is sanctioned by the dependency
between the source verb and the raising verb, which in WG is called ‘sharer’ (abbreviated ‘r’) because the raising verb shares its subject with the source verb. Thus if a word is the subject of a verb such as SEEM which also has a sharer, it must be the subject of the sharer as well.

Gerunds. Gerunds involve a mismatch of classification rather than of structure. They combine the characteristics of verbs with those of nouns, so they have is-a links to both: for example, not understanding syntax very well in sentence (12) is noun-like in its use as complement of a preposition and verb-like in its dependents.

(12) He worried us by not understanding syntax very well.

In contrast, nominalisations of verbs are simply nouns; for example, if understanding is a noun it may take determiners, adjectives and so on like a noun, but not direct objects like a verb (e.g. a deep understanding of syntax, but not *a deep understanding syntax). Multiple default inheritance allows verb and noun characteristics to combine very smoothly with very little additional attention in the grammar (Malouf 1998, Hudson 2000c). According to the grammar in Figure 27, understanding in (12) is the complement of by because it is a noun, and combines freely with not, syntax and very well because it is a verb. Notice that it is a noun but not a common noun, which explains why it does not combine freely with determiners or adjectives; and it is excluded from typical ‘verby’ positions because these all require a specific inflected form of a verb (finite or non-finite). DI plays an important part in this analysis by explaining exactly which properties gerunds inherit from each of the two super-categories, and also by allowing them to have a handful of exceptional properties such as possessive subjects.

Figure 27

Exceptional valency. This term is intended to cover the mismatches between default valency patterns associated with particular word classes and those of exceptional members of those word classes. For example, a typical English auxiliary verb takes a complement (the ‘sharer’ defined above) which is-a bare infinitive e.g. will come—but some verbs which in other respects are clearly auxiliary verbs have exceptional valency requirements: for those who still use these auxiliaries, OUGHT and USED take TO, and (for some speakers) HAVE10.

10 The auxiliary possessive HAVE is easy to recognise by its behaviour in interrogatives and negatives. If the interrogative of You have a car is Have you a car?, then have is an auxiliary verb because only auxiliary verbs can invert with their
takes an object noun (e.g. *I have a car*) (Hudson 1997). Figure 28 gives the grammar for this possessive HAVE in relation to the default pattern: by default an auxiliary takes a sharer, but exceptionally possessive HAVE takes an object. Such differences of valency can also produce mismatches within a word class between synonyms (Hudson et al 1996); for example, the adjectives LIKELY and PROBABLE appear to have exactly the same meaning and yet their valencies are different (compare *It is likely / 'probable to rain*).

Predicate nominals. Predicate nominals are an example of mismatch between syntax and semantics because noun phrases typically express arguments rather than predicates (Francis 1999). This is reflected in the very limited range of determiners allowed in such noun phrases — e.g. *She seems of / the / some nice person*. In WC terms, the contrast involves the familiar contrast between ‘sense’ and ‘referent’, and also the syntactic dependency ‘subject’. Any noun (indeed, almost any word) has a sense, which for a typical noun restricts the range of possible referents because the referent must be the sense; for example, the referent of *a dog is-a dog*. However a predicative noun exceptionally has a subject and limits this instead; for example, rather obviously, *Fido is a dog means 'Fido is-a dog'* (Hudson 1990:132). We can explain these facts by assuming a word-class ‘Predicative’ which regularly includes verbs and adjectives but only includes some nouns, and which exceptionally has a subject. The subject’s referent has some semantic relationship to the predicative’s sense, but exceptionally — in the case of predicative nouns — this relationship is ‘is-a’. Figure 29 shows how the default relation between a word’s sense and its referent (at the top) is replaced for predicative nouns by the relationship between its sense and its subject’s referent (in the bottom right corner).
Passives. Passives are the classic example of a mismatch between default and exceptional 'linking', whereby a verb's syntactic dependents are mapped onto the dependents of its sense—i.e., onto its arguments. In the WC analysis (Hudson 1989; Hudson 1990: 336-353), the passive verb's object doubles as its subject—i.e., its subject inherits from the same verb's object as well as from its subject—but the semantic role of the active subject is transferred to BY. Figure 30 shows how (though as in the other examples there is a great deal more to be said). The semantic roles 'er' and 'ee' are those typically assigned to the subject and object, but their content clearly varies from verb to verb (Hudson 1990: 157). If a verb's sense has an 'er', by default this is expressed by the referent of the verb's subject; and if it has an 'ee', this is expressed by the object's referent; but for passives these default pairings are overridden by a pairing of the 'er' with the referent of an adjunct (the by phrase) and by a merging of the subject and object dependents.

Figure 30

Re-classifiers. In spite of extensive discussion of words such as FAKE and FALSE, there does not seem to be an accepted cover term for them. The problem is that a fake diamond is not in fact a diamond, nor are false eye-lashes really eye-lashes, so their effect is to 're-classify' the things concerned. These words need a semantic analysis which shows that fake diamonds are intended to look like diamonds. To achieve this, we need to build on the meanings of other words such as the verbs FAKE (as in He faked the accident), COPY and SKETCH, all of which involve three participants: a person, a product and a model (which the product is meant to be similar to). Thus if John sketches Mary, he produces a sketch of her — i.e., a sketch for which she is the model. Once the relation 'model' is defined for these other meanings, we can use it in the analysis of fake diamond as something whose model is Diamond. This analysis has the advantage of showing that fake diamonds are at least meant to look like diamonds, even though they are not in fact diamonds. Figure 31 shows the default relationship between a word's sense and its dependent, whereby the dependent helps to define a hyponym of the sense; and it also shows the exceptional pattern for FAKE in which the default is-a relation is denied and replaced by the 'model' relation.\(^\text{12}\)

\(^{12}\)The crossed-out is-a link is a notational substitute for a proper analysis of negation. As the figure shows, a word in fact has a separate sense for each dependent to show how that dependent modifies its meaning; the effect is called 'semantic phrasing' (Hudson 1990: 146).
word's meaning, so the latter carries the meaning of the whole phrase. For example, in *long books about linguistics*, the word *books* means 'long books about linguistics' thanks to the modifying effect of the dependents. An *idiom* is an exceptional phrase whose head word carries a meaning which cannot be derived compositionally from its normal meaning plus that of its dependents—e.g., *hot dog* or *kick the bucket*. This means that the idiomatic use of the head word is stored as a special case of its lexeme whose meaning is stored ready-made and whose dependents are fixed; by DI, the characteristics of the special sub-case override those of the general lexeme. For example, *DOG/idiom* has a distinct sense (Hot-dog, defined as a roll containing a sausage) and always has the adjunct *hot*, which does not contribute to the meaning (a hot dog may in fact be cold), however relevant it may be to understanding the history of the idiom. Figure 32 shows this lexical entry.

![Figure 32](image)

**Figure 32**

*Pluralia tantum.* 'Pluralia tantum' is the traditional name for nouns such as *SCALES, SCISSORS and OATS* which are syntactically plural but semantically singular and which are included here as an example of the many mismatches between syntactic classification and semantics (It should be noted that such words are often translated by singular words in other languages; for example the German words for *SCISSORS and TROUSERS* are singular.). Such cases are generally accepted as evidence that syntax is distinct from semantics, though the evidence is sometimes disputed (Wierzbicka 1988). The default meaning of *Plural* is a set whose member is-a the sense (Hudson 1990:139), but this is overridden by the meanings of these lexical items as shown in Figure 33.

![Figure 33](image)

**Prototype effects.** The categories which serve as word senses have a strong tendency to show two kinds of prototype effects: degrees of goodness and degrees of membership. For example, Robin is better than Penguin as an example of Bird, and Chair is much more surely an example of Furniture than Computer is (Rosch 1977; Taylor 1995). Both of these effects are compatible with DI. Degree of goodness measures the number of exceptional features, while degree of membership involves the certainty of the is-a-relationship. We are sure that penguins are birds because we have been told so but nobody bothers to tell us whether or not computers are items of Furniture, so we have to work it out for ourselves. It is reasonable to assume that these effects are a general characteristic of how we categorize experience rather than a peculiarity of how we use language, but it should be reflected in any analysis of word meanings. For example, it should be clear that *Penguin* (the sense of *PENGUIN*) is-a Bird (the sense of *BIRD*), in spite of the non-default characteristics of penguins; and it should be unclear whether *Computer* is-a Furniture. Notice that this analysis does not require us to calculate and store the degree of fit for each potential member of a category, so
it escapes at least some criticisms of prototype theory (Fodor, Lepore 1996; Newmeyer 2000); prototype effects do not show that the category concerned is special—a 'prototype category'—nor that the category is defined in part by its 'degree of fit'. All they show is that some members override more default values than others do (Hudson 1984:40). Figure 34 suggests analyses for birds and furniture which show that—whereas we know that penguins are birds—we don’t know that computers are furniture, but we do know that they have at least two furniture-related characteristics (as well as two that qualify them as machines). These analyses are meant to be indicative rather than serious analyses.

![Diagram](image)

**Figure 34**

**Mistakes.** Finally, we can extend DI beyond the realms of grammar altogether and into mismatches in performance. The problem is to explain how we cope, as hearers and readers, with deviant input such as misspelt words. The fact is that we react to them in just the same way as to three-legged cats: we recognise them as deviant examples of recognisable categories rather than simply failing to classify them. If DI combined with the Best Fit Principle is indeed the logic of language processing as well as of the rest of life, this is easy to explain. Figure 35 illustrates how we recognise *yellow* as a misspelling of *YELLOW* and how we then exploit this classification by inheriting the latter’s unobservable characteristics including its meaning. The crucial part of the diagram is the *is-a* link from the input *yellow* to {*yellow*}; as usual, an *is-a* link does not require total identity but allows overriding.

![Diagram](image)

**Figure 35**

This section has surveyed a wide range of mismatch phenomena in language to which DI can plausibly be applied. It is obviously not intended to prove that the proposed analyses are correct; in every case this would require a great deal more detailed evidence and discussion. The aim has been to support my general contention that DI lies behind a very wide range of mismatches, and possibly even behind every type of mismatch without exception.
6 The architecture of language

I should like to finish with some general observations on the relation between mismatches and the architecture of language. It is true that mismatches can be used as evidence for the separateness of levels; I mentioned the case of pluralia tantum which show that semantic plurality is different from syntactic plurality, and I fully accept the logic. However it seems to me that mismatches also undermine the standard View of levels; I mentioned the case of pluralia tantum which show that semantic plurality is different from syntactic plurality, and I fully accept the logic. However it seems to me that mismatches also undermine the standard View of levels (at least, the one found outside Chomskyan circles), whereby each level has its own distinct vocabulary of units and its own internal organisation. This view is largely valid and helpful in straightforward cases, but the levels are less watertight than is often claimed.

Figure 36 presents a fairly standard View of language, in which a number of distinct levels of representation are related to each other by interface relations (Jackendoff's 'correspondence rules' Jackendoff 1997). Each level (except perhaps the semantic level) defines a distinct set of elements—words, morphemes, phonemes—which combine only with one another; and each level is mapped onto the adjacent levels by relations which are equally selective. For example, 'meaning' relates a word to a semantic structure while 'form' relates it to a morphological one.

Figure 36

It is true that some parts of Figure 36 are controversial; for example, the whole of syntax is presented as a collection of more or less general 'lexical entries' (understood as entries for word-size categories), so there is no distinction between a lexicon and a set of rules. However the crucial characteristic of syntax is that its basic building blocks are words, and everybody agrees on this, so we can ignore the relatively unimportant question of how word combinations should be represented. Figure 36 also gives a higher profile to the level of morphology than in many other theories; but I think it would be fair to say that everyone recognises morphology as distinct at least from phonology and probably also from syntax, so it seems quite reasonable to include it as a distinct level.

![Diagram of language levels](image)

This model of language works in straightforward cases. For example, consider the sentence Joe loves carrots, whose structure is sketched in Figure 37. The word loves is combined with two other words and relates to the concept 'Joe loving carrots' and to the morphemes {love} and {carrots}.

Figure 37

14 The division between 'lexicon' and 'rules' is both unclear and contentious. Most theories of language structure take it for granted, but 'cognitive' theories reject it. This is true not only of Word Grammar (Hudson 1990:33) but also of Cognitive Grammar (Langacker 2000) and Construction Grammar (Goldberg 1998).
which as predicted combine with one another and individually map to a phonological structure. In such examples, the traditional model works well. However, when we look for exceptions, we find them at every point in the model: combinations of units which do not all belong to the same level and which therefore infringe both the inter-level and intra-level constraints. The exceptions are summarised in Table 2, and explained a little more fully below.

<table>
<thead>
<tr>
<th>Table 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>level(s)</td>
</tr>
<tr>
<td>semantics/syntax</td>
</tr>
<tr>
<td>syntax</td>
</tr>
<tr>
<td>syntax/morphology</td>
</tr>
<tr>
<td>morphology</td>
</tr>
<tr>
<td>morphology/phonology</td>
</tr>
</tbody>
</table>

Metalanguage. The default meaning is not part of language, but exceptionally some vocabulary is dedicated to referring to linguistic concepts (e.g. WORD, NOUN) (Hudson 1990:73-74). Words can also be used to refer to themselves (e.g. 'But should not start a sentence').

Non-linguistic complements. By default words combine with other words, but there are a few words which require a complement which is not a word, and which is not even part of language (Hudson 1989, Hudson 1990: 67). This is true of the verb GO, in sentences such as The train went [train-noise]. Some kind of noise or action is obligatory, so it is part of the verb's syntactic valency, but it must not be a word or word-string—e.g. *He went 'It's for you.' is not grammatical for me (though it may be for my children's generation).

Clitics. As explained earlier, a clitic is a word which is part of another word—a clear exception to the principle that word-parts are morphemes.

Cranberry morphs. Typically morphemes combine only with other morphemes, but structuralist morphology recognised that words such as CRANBERRY are problematic because they contain a morpheme (berry)—with a residue which does not qualify as a morpheme by the usual principles. A reasonable conclusion is that the residue is not a morpheme but simply a phonological pattern.

Writing. Writing is a rather obvious exception to the principle that morphemes consist of phonemes; for example, the morphemes {love} and [Z] are usually written <love> and <<>. However good the reasons may be for giving priority to speech as the default medium for language, writing must be integrated into the model of language because it is indubitably part of the same cognitive system (e.g. those who can read can also read aloud; and spelling sometimes influences pronunciation).

Sub-phonemic detail. The standard view of phonology recognises a limited range of phoneme-like units out of which all morphemes are built, with sub-phonemic detail being supplied by general context-sensitive rules. However, there is considerable evidence that some sub-phonemic detail can be stored with lexical items (Bybee 2000). Similarly we regularly store sub-phonemic detail with loan words; for example, my stored entry for RESTAURANT must specify the non-English nasal vowel that I use alongside a non-French /r/.

In conclusion, then, the levels of language 'leak', and in some cases the leaching is quite severe. This means not that the standard model is wrong, but that it only describes the most typical patterns of language. As a research strategy, it is sensible to focus first on typical patterns, but it would be wrong to ignore other patterns if we want a complete understanding of language which includes the mismatches as well as the defaults.

References


References


12

Cooperating Constructions

ROBERT MALOUF

1 Introduction

Optimality theoretic (OT) syntax (Dekkers et al. 2000, Legendre et al. 2000, Sells 2001) is an important development in the world of constraint based grammar, because it introduces a new way of looking at constraint interactions. The leading idea is that not all constraints are "surface true" - constraints may be violated when they have to be in order to satisfy some more important, higher ranking constraint. This opens up an interesting new way of looking at mismatch phenomena. A mismatch arises when a constraint imposing 'match', a correspondence between grammatical modules or levels, must be violated in order to satisfy some higher ranking module's internal constraints.

OT also suggests a natural way for integrating functional motivations into formal theories of grammar. If the constraints are appropriately grounded in functional considerations, then an OT grammar is the consequence of finding a balance among various competing communicative drives. Cases of mismatch that arise out of this competition can be seen as a direct outcome of conflicting functional pressures.

Unfortunately, however, the fit between OT and constraint based grammar is not entirely comfortable. Informally, one can sketch the logic of OT constraint interaction as follows (Prince and Smolensky 1993). First, start with a partial representation of an utterance (say, something like logical form). From this partial representation, generate the complete (usually infinite) set of possible complete representations that can be built from the input. Next, evaluate each of these candidate representations with respect to the highest ranked constraint, and

Mismatch: Form Function Incompatibility and the Architecture of Grammar.
Paine J. Francis and Laura A. Michaelis (eds.).
Copyright © 2003, CSLI Publications.