Word Grammar[1]

Richard Hudson

Contents

1 A brief overview of the theory

Word Grammar (**WG**) is a general theory of language structure. Most of the work to date has dealt with syntax, but there has also been serious work in semantics and some more tentative explorations of morphology, sociolinguistics, historical linguistics and language processing. The only areas of linguistics that have not been addressed at all are phonology and language acquisition (but even here see van Langendonck 1987). The aim of this article is breadth rather than depth, in the hope of showing how far-reaching the theory's tenets are.

Although the roots of WG lie firmly in linguistics, and more specifically in grammar, it can also be seen as a contribution to cognitive psychology; in terms of a widely used classification of linguistic theories, it is a branch of *cognitive linguistics* (Taylor 1989, Lakoff 1987, Langacker 1987; 1990). The theory has been developed from the start with the aim of integrating all aspects of language into a single theory which is also compatible with what is know about general cognition. This may turn out not to be possible, but to the extent that it is possible it will have explained the general characteristics of language as `merely' one instantiation of more general cognitive characteristics.

The overriding consideration, of course, is the same as for any other linguistic theory: to be true to the facts of language structure. However, our assumptions make a great deal of difference when approaching these facts, so it is possible to arrive at radically different analyses according to whether we assume that language is a unique module of the mind, or that it is similar to other parts of cognition. The WG assumption is that language can be analysed and explained in the same way as other kinds of knowledge or behaviour unless there is clear evidence to the contrary. So far this strategy has proved productive and largely successful, as we shall see below.

1 This paper was originally written in 1998 for publication in *Dependency and Valency. An International Handbook of Contemporary Research* (edited by Vilmos Ágel; Ludwig M. Eichinger; Hans-Werner Eroms; Peter Hellwig; Hans Jürgen Heringer; and Henning Lobin) Berlin: Walter de Gruyter. It is unclear whether this handbook will ever appear, so it is good to have the opportunity to publish a revised version here.

As the theory's name suggests, the central unit of analysis is the **word**, which is central to all kinds of analysis:

- Grammar. Words are the only units of syntax (section 8), as sentence structure consists entirely of dependencies between individual words; WG is thus clearly part of the tradition of *dependency grammar* dating from Tesnière (1959; Fraser 1994). Phrases are implicit in the dependencies, but play no part in the grammar. Moreover, words are not only the largest units of syntax, but also the smallest. In contrast with Chomskyan linguistics, syntactic structures do not, and cannot, separate stems and inflections, so WG is an example of *morphology-free syntax* (Zwicky 1992, 354). Unlike syntax, morphology (section 7) is based on constituent-structure, and the two kinds of structure are different in others ways too.
- Semantics. As in other theories words are also the basic lexical units where sound meets syntax and semantics, but in the absence of phrases words also provide the only point of contact between syntax and semantics, giving a radically `lexical' semantics. As will appear in section 9, a rather unexpected effect of basing semantic structure on single words is a kind of phrase structure in the semantics.
- Situation. We shall see in section 6 that words are the basic units for contextual analysis (in terms of deictic semantics, discourse or sociolinguistics).

Words, in short, are the nodes that hold the `language' part of the human network together. This is illustrated by the word <u>cycled</u> in the sentence <u>I cycled to UCL</u>, which is diagrammed in Figure 1.

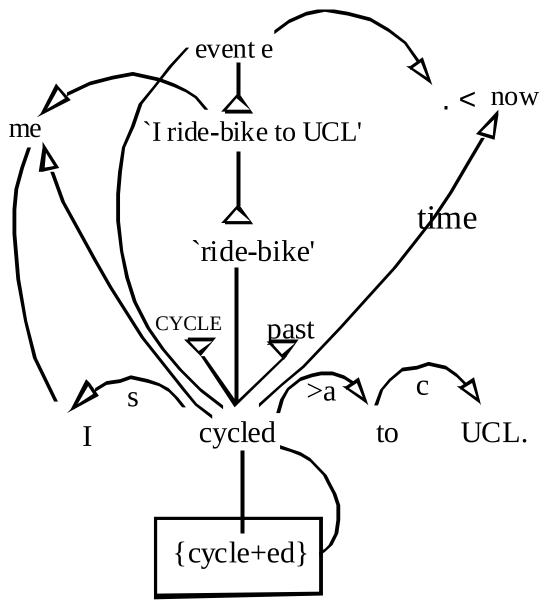


Figure 1

As can be seen in this diagram, <u>cycled</u> is the meeting point for ten relationships which are detailed in Table 1. These relationships are all quite traditional (syntactic, morphological, semantic, lexical and contextual), and traditional names are used where they exist, but the diagram uses notation which is peculiar to WG. It should be easy to imagine how such relationships can multiply to produce a rich network in which words are related to one another as well as to other kinds of element including morphemes and various kinds of meaning. All these elements, including the words themselves, are **`concepts'** in the standard sense; thus a WG diagram is an attempt to model a small part of the total conceptual network of a typical speaker.

<table>
<tr><th>related concept C</th><th>relationship of C to cycled</th><th>notation in diagram</th></tr>
<tr><td>the word I</td><td>subject</td><td>`s'</td></tr>
<tr><td>the word to</td><td>post-adjunct</td><td>`>a'</td></tr>
<tr><td>the morpheme {cycle}</td><td>stem</td><td>straight downward line</td></tr>
<tr><td>the word-form {cycle+ed}</td><td>whole</td><td>curved downward line</td></tr>
<tr><td>the concept `ride-bike'</td><td>sense</td><td>straight upward line</td></tr>
<tr><td>the concept `event e'</td><td>referent</td><td>curved upward line</td></tr>
<tr><td>the lexeme CYCLE</td><td rowspan="2">cycled isa C</td><td rowspan="2">triangle resting on C</td></tr>
<tr><td>the inflection 'past'</td></tr>
<tr><td>me</td><td>speaker</td><td>`speaker'</td></tr>
<tr><td>now</td><td>time</td><td>`time'</td></tr>
</table>

Table 1

2 Historical background

The theory described in this article is the latest in a family of theories which have been called `Word Grammar' since the early 1980s (Hudson 1984). The present theory is very different in some respects from the earliest one, but the continued use of the same name is justified because we have preserved some of the most fundamental ideas - the central place of the word, the idea that language is a network, the role of default inheritance, the clear separation of syntax and semantics, the integration of sentence and utterance structure. The theory is still changing and a range of more or less radical variants are under active consideration (Rosta 1994; 1996; 1997, Kreps 1997). The version that I shall present here is the one that I myself find most coherent and convincing as I write in 2002.

As in other theories, the changes have been driven by various forces - new data, new ideas, new alternative theories, new personal interests; and by the influence of teachers, colleagues and students. The following brief history may be helpful in showing how the ideas that are now called `Word Grammar' developed during my academic life.

The 1960s. My PhD analysis of Beja used the theory being developed by Halliday (1961) under the name `Scale-and-Category' grammar, which later turned into Systemic Functional Grammar (Halliday 1985, Butler 1985). I spent the next six years working with Halliday, whose brilliantly wide-ranging analyses impressed me a lot. Under the influence of Chomsky's generative grammar (1957, 1965), reinterpreted by McCawley (1968) as well-formedness conditions, I developed the first generative version of Halliday's Systemic Grammar (Hudson 1970). This theory has a very large network (the `system network') at its heart, and networks also loomed large at that time in the Stratificational Grammar of Lamb

(1966, Bennett 1994). Another reason why stratificational grammar was important was that it aimed to be a model of human language processing - a cognitive model.

The 1970s. Seeing the attractions of both valency theory and Chomsky's subcategorisation, I produced a hybrid theory which was basically Systemic Grammar, but with the addition of word-word dependencies under the influence of Anderson (1971); the theory was called `Daughter-Dependency Grammar' (Hudson 1976). Meanwhile I was teaching sociolinguistics and becoming increasingly interested in cognitive science (especially **default inheritance** systems and frames) and the closely related field of lexical semantics (especially Fillmore's Frame Semantics 1975, 1976). The result was a very `cognitive' textbook on sociolinguistics (Hudson 1980a, 1996a). I was also deeply influenced by Chomsky's `Remarks on nominalisation' paper (1970), and in exploring the possibilities of a radically lexicalist approach I toyed with the idea of `pan-lexicalism' (1980b, 1981): everything in the grammar is `lexical' in the sense that it is tied to word-sized units (including word classes).

The 1980s. All these influences combined in the first version of Word Grammar (Hudson 1984), a cognitive theory of language as a **network** which contains both `the grammar' and `the lexicon' and which integrates language with the rest of cognition. The semantics follows Lyons (1977), Halliday (1967-8) and Fillmore (1976) rather than formal logic, but even more controversially, the syntax no longer uses phrase structure at all in describing sentence structure, because everything that needs to be said can be said in terms of **dependencies** between single words. The influence of continental dependency theory is evident but the dependency structures were richer than those allowed in `classical' dependency grammar (Robinson 1970) - more like the functional structures of Lexical Functional Grammar (Kaplan and Bresnan 1982). Bresnan's earlier argument (1978) that grammar should be compatible with a psychologically plausible parser also suggested the need for a parsing algorithm, which has led to a number of modest NLP systems using WG (Fraser 1985; 1989; 1993, Hudson 1989, Shaumyan 1995). These developments provided the basis for the next book-length description of WG, `English Word Grammar' (**EWG**, Hudson 1990). This attempts to provide a formal basis for the theory as well as a detailed application to large areas of English morphology, syntax and semantics.

The 1990s. Since the publication of EWG there have been some important changes in the theory, ranging from the general theory of default inheritance, through matters of syntactic theory (with the addition of `surface structure', the virtual abolition of features and the acceptance of 'unreal' words) and morphological theory (where 'shape', `whole' and `inflection' are new), to details of analysis, terminology and notation. These changes will be described below. WG has also been applied to a wider range of topics than previously:

- lexical semantics (Gisborne 1993, 1996, 2000, 2001, Hudson and Holmes 2003, Hudson 1992, 1995, 2003a, Sugayama 1993, 1996, 1998),
- morphology (Creider 1999, Creider and Hudson 1999),
- historical linguistics (Hudson 1997a, b),
- sociolinguistics (Hudson 1996a; 1997b),
- language processing (Hudson 1993a, b; 1996b; Hiranuma 1999, 2001).

Most of the work done since the start of WG has applied the theory to English, but it has also been applied to the following languages: Tunisian Arabic (Chekili 1982), Greek (Tsanidaki 1995, 1996a, b), Italian (Volino 1990), Japanese (Sugayama 1991, 1992, 1993, 1996, Hiranuma 1999, 2001) and Polish (Gorayska 1985).

The theory continues to evolve, and at the time of writing a `Word Grammar

Encyclopedia' which can be downloaded via the WG web-site (http://www.phon.ucl.ac.uk/home/dick/wg.htm) is updated in alternate years.

3 The cognitive network

3.1 Language as part of a general network

The basis for WG is an idea which is quite uncontroversial in cognitive science:

> The idea is that memory connections provide the basic building blocks through which our knowledge is represented in memory. For example, you obviously know your mother's name; this fact is recorded in your memory. The proposal to be considered is that this memory is literally represented by a memory connection, ... That connection isn't some appendage to the memory. Instead, the connection is the memory. ... all of knowledge is represented via a sprawling network of these connections, a vast set of associations. (Reisberg 1997:257-8)

In short, knowledge is held in memory as an **associative network**. What is more controversial is that, according to WG, the same is true of our knowledge of words, so the sub-network responsible for words is just a part of the total `vast set of associations'. Our knowledge of words is our language, so our language is a network of associations which is closely integrated with the rest of our knowledge.

However uncontroversial (and obvious) this view of knowledge may be in general, it is very controversial in relation to language. The only part of language which is widely viewed as a network is the lexicon (Aitchison 1987:72), and a fashionable view is that even here only lexical irregularities are stored in an associative network, in contrast with regularities which are stored in a fundamentally different way, as `rules' (Pinker and Prince 1988). For example, we have a network which shows for the verb <u>come</u> not only that its meaning is `come' but that its past tense is the irregular <u>came</u>, whereas regular past tenses are handled by a general rule and not stored in the network. The WG view is that exceptional and general patterns are indeed different, but they can both be accommodated in the same network because it is an `inheritance network' in which general patterns and their exceptions are related by default inheritance (which is discussed in more detail in section 4). To pursue the last example, both patterns can be expressed in exactly the same prose:

(1) The shape of the past tense of a verb consists of its stem followed by <u>-d</u>.

(2) The shape of the past tense of <u>come</u> consists of <u>came</u>.

The only difference between these rules lies in two places: `a verb' versus <u>come</u>, and `its stem followed by <u>-ed</u>' versus <u>came</u>. Similarly, they can both be incorporated into the same network, as shown in Figure 2 (where the triangle once again shows the `isa' relationship by linking the general concept at its base to the specific example connected to its apex).

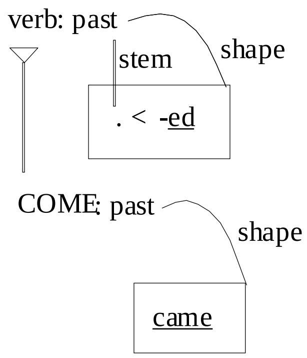


Figure 2

Once the possibility is accepted that some generalisations may be expressed in a network, it is easy to extend the same treatment to the whole grammar, as we shall see in later examples. One consequence, of course, is that we lose the formal distinction between `the lexicon' and `the rules' (or `the grammar'), but this conclusion is also accepted outside WG in Cognitive Grammar (Langacker 1987) and Construction Grammar (Goldberg 1995). The only parts of linguistic analysis that cannot be included in the network are the few general theoretical principles (such as the principle of default inheritance).

3.2 Labelled links

It is easy to misunderstand the network view because (in cognitive psychology) there is a long tradition of `associative network' theories in which all links have just the same status: simple `association'. This is not the WG view, nor is it the view of any of the other theories mentioned above, because links are **classified** and labelled - `stem', `shape', `sense', `referent', `subject', `adjunct' and so on and on. The classifying categories range from the most general - the `isa' link - to categories which may be specific to a handful of concepts, such as `goods' in the framework of commercial transactions (Hudson 2003a). This is a far cry from the idea of a network of mere `associations' (such as underlies connectionist models). One of the immediate benefits of this approach is that it allows named links to be used as functions, in the mathematical sense of Kaplan and Bresnan (1982, 182), which yield a unique value - e.g. `the referent of the subject of the verb' defines one unique concept for each verb. In order to distinguish this approach from the traditional associative networks we can call these networks `labelled'.

Even within linguistics, labelled networks are controversial because the labels themselves need an explanation or analysis. Because of this problem some theories avoid labelled relationships, or reduce labelling to something more primitive: for example, Chomsky has always avoided functional labels for constituents such as `subject' by using configurational definitions, and the predicate calculus avoids semantic role labels by distinguishing arguments in terms of order.

There is no doubt that labels on links are puzzlingly different from the labels that we give to the concepts that they link. Take the small network in Figure 2 for past tenses. One of the nodes is labelled `COME: past', but this label could in fact be removed without any effect because `COME: past' is the only concept which isa `verb: past' and which has came as its shape. Every concept is uniquely defined by its links to other concepts, so labels are redundant (Lamb 1996, 1999:59). But the same is not true of the labels on links, because a network with unlabelled links is a mere associative network which would be useless in analysis. For example, it is no help to know that in John saw Mary the verb is linked, in some way or other, to the two nouns and that its meaning is linked, again in unspecified ways, to the concepts `John' and `Mary'; we need to know which noun is the subject, and which person is the see-er. The same label may be found on many different links - for example, every word that has a sense (i.e. virtually every word) has a link labelled `sense', every verb that has a subject has a `subject' link, and so on. Therefore the function of the labels is to classify the links as same or different, so if we remove the label we lose information. It makes no difference whether we show these similarities and differences by means of verbal labels (e.g. `sense') or some other notational device (e.g. straight upwards lines); all that counts is whether or not our notation classifies links as same or different. Figure 3 shows how this can be done using first conventional attribute-value matrices and second, the WG notation used so far.

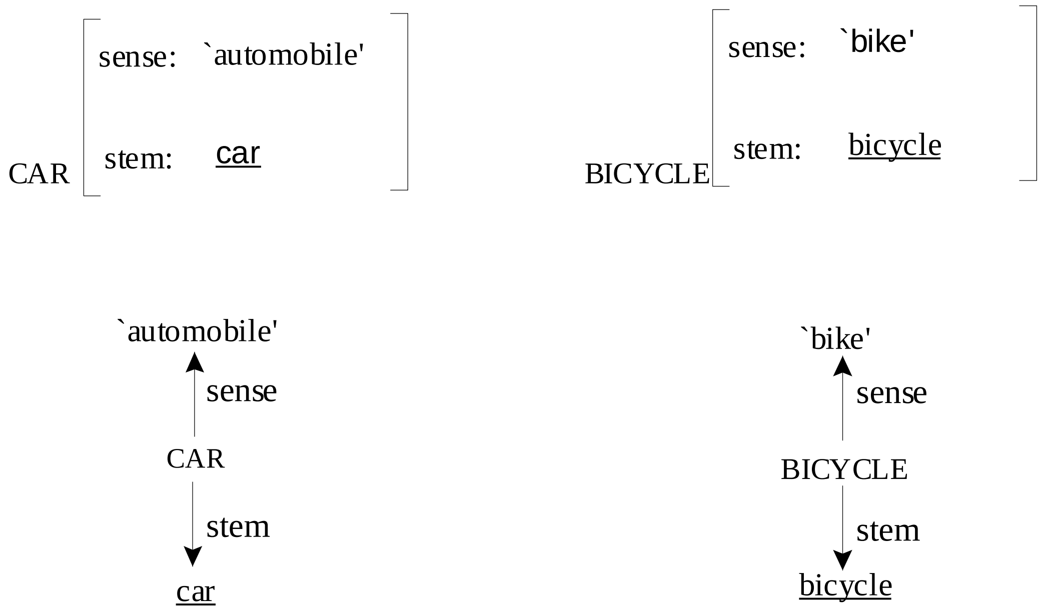


Figure 3

This peculiarity of the labels on links brings us to an important characteristic of the network approach which allows the links themselves to be treated like the concepts which they link - as `second-order concepts', in fact. The essence of a network is that each concept should be represented just once, and its multiple links to other concepts should be shown as multiple links, not as multiple copies of the concept itself. Although the same principle applies generally to attribute-value matrices, it does not apply to the attributes themselves. Thus there is a single matrix for each concept, and if two attributes have the same value this is shown (at least in one notation) by an arc that connects the two value-slots. But when it comes to the attributes themselves, their labels are repeated across matrices (or even within a single complex matrix). For example, the matrix for a raising verb contains within it the matrix for its complement verb; an arc can show that the two subject slots share the same filler but the only way to show that these two slots belong to the same attribute is to repeat the label `subject'.

In a network approach it is possible to show both kinds of identity in the same way: by means of a single node with multiple `isa' links. If two words are both nouns, we show this by an isa link from each to the concept `noun'; and if two links are both `subject' links, we put an isa link from each link to a single general `subject' link. Thus labelled links and other notational tricks are just abbreviations for a more complex diagram with second-order links between links. These second-order links are illustrated in Figure 4 for car and bicycle, as well as for the sentence Jo snores.

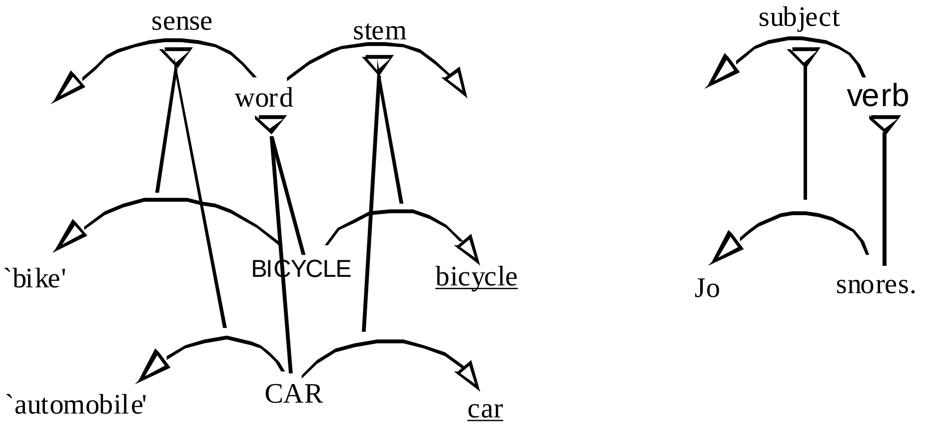


Figure 4

This kind of analysis is too cumbersome to present explicitly in most diagrams, but it is important to be clear that it underlies the usual notation because it allows the kind of analysis which we apply to ordinary concepts to be extended to the links between them. If ordinary concepts can be grouped into larger classes, so can links; if ordinary concepts can be learned, so can links. And if the labels on ordinary concepts are just mnemonics which could, in principle, be removed, the same is true of the labels on all links except the `isa' relationship itself, which reflects its fundamental character.

3.3 Modularity

The view of language as a labelled network has interesting consequences for the debate about modularity: is there a distinct `module' of the mind dedicated exclusively to language (or to some part of language such as syntax or inflectional morphology)? Presumably not if a module is defined as a separate `part' of our mind and if the language network is just a small part of a much larger network. One alternative to this strong version of modularity is no modularity at all, with the mind viewed as a single undifferentiated whole; this seems just as wrong as a really strict version of modularity. However there is a third possibility. If we focus on the links, any such network is inevitably `modular' in the much weaker (and less controversial) sense that links between concepts tend to cluster into relatively dense **sub-networks** separated by relatively sparse boundary areas.

Perhaps the clearest evidence for some kind of modularity comes from language pathology, where abilities are impaired selectively. Take the case of Pure Word Deafness (Altman 1997:186), for example. Why should a person be able to speak and read normally, and to hear and classify ordinary noises, but not be able to understand the speech of other people? In terms of a WG network, this looks like an inability to follow one particular link-type (`sense') in one particular direction (from word to sense). Whatever the reason for this strange disability, at least the WG analysis suggests how it might apply to just this one aspect of language, while also applying to every single word: what is damaged is the general relationship `sense', from which all particular sense relationships are inherited. A different kind of problem is illustrated by patients who can name everything except one category - e.g. body-parts or things typically found indoors (Pinker 1994:314). Orthodox views on modularity seem to be of little help in such cases, but a network approach at least explains how the non-linguistic concepts concerned could form a mental cluster of closely-linked and mutually defining concepts with a single super-category. It is easy to imagine reasons why such a cluster of concepts might be impaired selectively (e.g. that closely related concepts are stored close to each other, so a single injury could sever all their sense links), but the main point is to have provided a way of unifying them in preparation for the explanation.

In short, a network with classified relations allows an injury to apply to specific relation types so that these relations are disabled across the board. The approach also allows damage to specific areas of language which form clusters with strong internal links and weak external links. Any such cluster or shared linkage defines a kind of `module' which may be impaired selectively, but the module need not be innate: it may be `emergent', a cognitive pattern which emerges through experience (Bates et al 1998, Karmiloff-Smith 1992).

4 Default inheritance

Default inheritance is just a formal version of the logic that linguists have always used: true generalisations may have exceptions. We allow ourselves to say that verbs form their past tense by adding -ed to the stem even if some verbs don't, because the specific provision made for these exceptional cases will automatically override the general pattern. In short, characteristics of a general category are `inherited' by instances of that category only `by default' - only if they are not overridden by a known characteristic of the specific case. Common sense tells us that this is how ordinary inference works, but default inheritance only works when used sensibly. Although it is widely used in artificial intelligence, researchers treat it with great caution (Luger and Stubblefield 1993:386-8). The classic formal treatment is Touretsky (1986).

Inheritance is carried by the `isa' relation, which is another reason for considering this relation to be fundamental. For example, because snores *isa* `verb' it automatically inherits all the known characteristics of `verb' (i.e. of `the typical verb'), including for example the fact that it has a subject; similarly, because the link between Jo and snores in Jo snores isa `subject' it inherits the characteristics of `subject'. As we have already seen, the notation for `isa' consists of a small triangle with a line from its apex to the instance. The base of the triangle which rests on the general category reminds us that this category is larger than the instance, but it can also be imagined as the mouth of a hopper into which information is poured so that it can flow along the link to the instance.

The mechanism whereby default values are overridden has changed during the last few years. In EWG, and also in Fraser and Hudson (1992), the mechanism was `stipulated overriding', a system peculiar to WG; but since then this system has been abandoned. WG now uses a conventional system in which a fact is automatically blocked by any other fact which conflicts and is more specific. Thus the fact that the past tense of COME is came automatically blocks the inheritance of the default pattern for past tense verbs. One of the advantages of a network notation is that this is easy to define formally: we always prefer the value for `R of C' (where R is some relationship, possibly complex, and C is a concept) which is nearest to C (in terms of intervening links). For example, if we want to find the shape of the past tense of COME, we have a choice between came and comed, but the route to came is shorter than that to comed because the latter passes through the concept `past tense of a verb'. (For detailed discussions of default inheritance in WG, see Hudson 2000a, 2003b.)

Probably the most important question for any system that uses default inheritance concerns **multiple inheritance**, in which one concept inherits from two different concepts simultaneously - as `dog' inherits, for example, both from `mammal' and from `pet'. Multiple inheritance is allowed in WG, as in unification-based systems and the programming language DATR (Evans and Gazdar 1996); it is true that it opens up the possibility of conflicting information being inherited, but this is a problem only if the conflict is an artefact of the analysis. There seem to be some examples in language where a form is ungrammatical precisely because there is an irresoluble conflict between two characteristics; for example, in many varieties of standard English the combination *I amn't is predictable, but ungrammatical. One explanation for this strange gap is that the putative form amn't has to inherit simultaneously from aren't (the negative present of BE) and am (the I-form of BE); but these models offer conflicting shapes (aren't, am) without any way for either to override the other (Hudson 2000). In short, WG does allow multiple inheritance, and indeed uses it a great deal (as we shall see in later sections).

5 The language network

According to WG, then, language is a network of concepts. The following more specific claims flesh out this general idea.

First, language is part of the same general conceptual network which contains many concepts which are not part of language. What distinguishes the language area of this network from the rest is that the concepts concerned are words and their immediate characteristics. This is simply a matter of definition: concepts which are not directly related to words would not be considered to be part of language. As explained in section 3.3, language probably qualifies as a module in the weak sense that the links among words are denser than those between words and other kinds of concept, but this does not mean that language is a module in the stronger sense of being `encapsulated' or having its own special formal characteristics. This is still a matter of debate, but we can be sure that at least some of the characteristics of language are also found elsewhere - the mechanism of default inheritance and the isa relation, the notion of linear order, and many other formal properties and principles.

As we saw in Table 1, words may have a variety of links to each other and to other concepts. This is uncontroversial, and so are most of the links that are recognised. Even the traditional notions of `levels of language' are respected in as much as each level is defined by a distinct kind of link: a word is linked to its morphological structure via the `stem' and `shape' links, to its semantics by the `sense' and `referent' links, and to its syntax by dependencies and word classes. Figure 5 shows how clearly the traditional levels can be separated from one another. In WG there is total commitment to the `autonomy' of levels, in the sense that the levels are formally distinct.

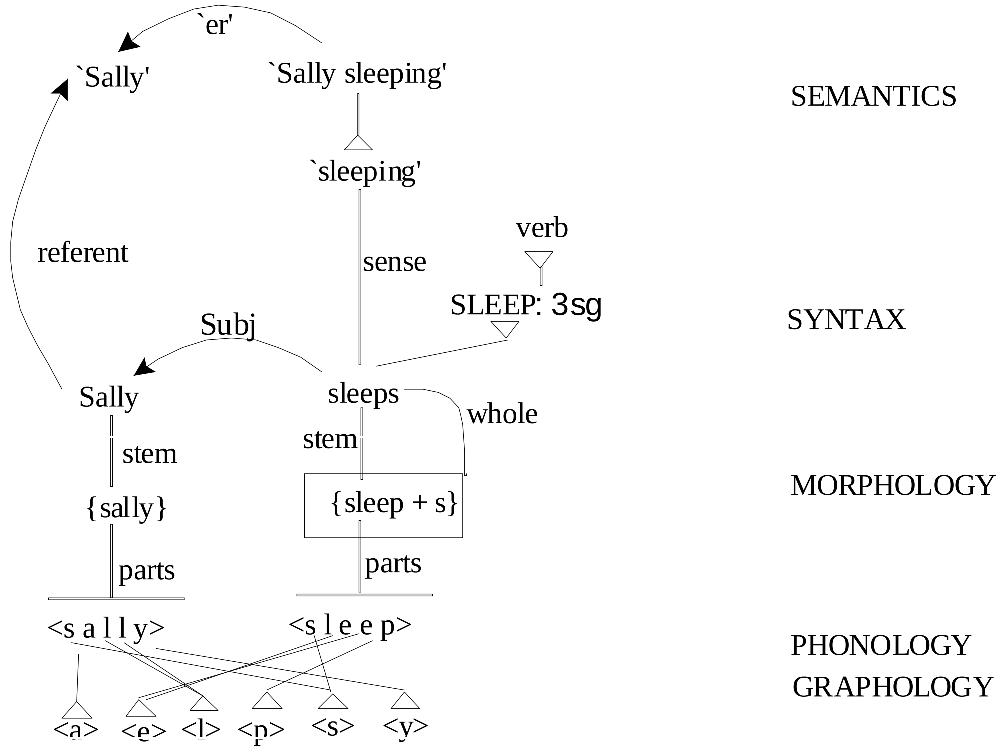


Figure 5

The most controversial characteristic of WG, at this level of generality, is probably the central role played by **inheritance (isa) hierarchies**. Inheritance hierarchies are the sole means available for classifying concepts, which means that there is no place for feature-descriptions. In most other theories, feature-descriptions are used to name concepts, so that instead of `verb' we have `[+V, -N]' or (changing notation) `[Verb:+, Noun:-, SUBCAT:<NP>]' or even `S/NP'. This is a fundamental difference because, as we saw earlier, the labels on WG nodes are simply mnemonics and the analysis would not be changed at all if they were all removed. The same is clearly not true where feature-descriptions are used, as the name itself contains crucial information which is not shown elsewhere. To classify a word as a verb in WG we give it an isa link to `verb'; we do not give it a feature-description which contains that of `verb'.

The most obviously classifiable elements in language are words, so in addition to specific, unique, words we recognise general `word-types'; but we can refer to both simply as `words' because (as we shall see in the next section) their status is just the same. Multiple inheritance allows words to be classified on two different `dimensions': as lexemes (DOG, LIKE, IF, etc) and as inflections (plural, past, etc). Figure 6 shows how this cross-classification can be incorporated into an isa hierarchy. The traditional word classes are shown on the lexeme dimension as classifications of lexemes, but they interact in complex ways with inflections. Cross-classification is possible even among word-classes; for example, English gerunds (e.g. Writing in Writing articles is fun.) are both nouns and verbs (Hudson 2000b), and in many languages participles are probably both adjectives and verbs.

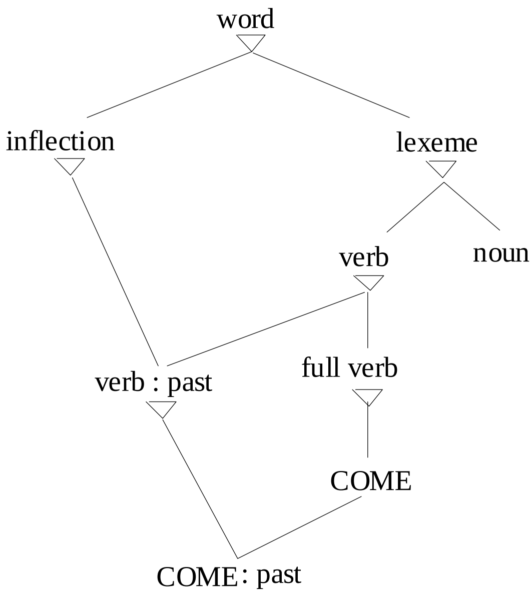


Figure 6

Unlike other theories, the classification does not take words as the highest category of concepts - indeed, it cannot do so if language is part of a larger network. WG allows us to show the similarities between words and other kinds of communicative behaviour by virtue of an isa link from `word' to `communication', and similar links show that words are actions and events. This is important in the analysis of deictic meanings which have to relate to the participants and circumstances of the word as an action.

This hierarchy of words is not the only isa hierarchy in language. There are two more for speech sounds (`phonemes') and for letters ('graphemes'), and a fourth for morphemes and larger 'forms' (Hudson 1997b, Creider and Hudson 1999), but most important is the one for relationships - `sense', `subject' and so on. Some of these relationships belong to the hierarchy of dependents which we shall discuss in the section on syntax, but there are many others which do not seem to comprise a single coherent hierarchy peculiar to language (in contrast with the `word' hierarchy). What seems much more likely is that relationships needed in other areas of thought (e.g. `before', `part-of') are put to use in language.

To summarise, the language network is a collection of words and word-parts (speech-sounds, letters and morphemes) which are linked to each other and to the rest of cognition in a variety of ways, of which the most important is the `isa' relationship which classifies them and allows default inheritance.

6 The utterance network

A WG analysis of an **utterance** is also a network; in fact, it is simply an extension of the permanent cognitive network in which the relevant word tokens comprise a `fringe' of temporary concepts attached by `isa' links, so the utterance network has just the same formal characteristics as the permanent network. For example, suppose you say to me `I agree.' My task, as hearer, is to segment your utterance into the two words I and agree, and then to classify each of these as an example of some word in my permanent network (my grammar). This is possible to the extent that default inheritance can apply smoothly; so, for example, if my grammar says that I must be the subject of a tensed verb, the same must be true of this token, though as we shall see below, exceptions can be tolerated. In short, a WG grammar can generate representations of actual utterances, warts and all, in contrast with most other kinds of grammar which generate only idealised utterances or `sentences'. This blurring of the boundary between grammar and utterance is very controversial, but it follows inevitably from the cognitive orientation of WG.

The status of utterances has a number of theoretical consequences both for the structures generated and for the grammar that generates them. The most obvious consequence is that word tokens must have different names from the types of which they are tokens; in our example, the first word must not be shown as I if this is also used as the name for the word-type in the grammar. This follows from the fact that identical labels imply identity of concept, whereas tokens and types are clearly distinct concepts. The WG convention is to reserve conventional names for types, with tokens labelled `w1', `w2' and so on through the utterance. Thus our example consists of w1 and w2, which isa `I' and `AGREE:pres' respectively. This system allows two tokens of the same type to be distinguished; so in I agree I made a mistake, w1 and w3 both isa `I'. (For simplicity WG diagrams in this paper only respect this convention when it is important to distinguish tokens from types.)

Another consequence of integrating utterances into the grammar is that word types and tokens must have characteristics such that a token can inherit them from its type. Obviously the token must have the familiar characteristics of types - it must belong to a lexeme and a word class, it must have a sense and a stem, and so on. But the implication goes in the other direction as well: the type may mention some of the token's characteristics that are normally excluded from grammar, such as characteristics of the speaker, the addressee and the situation. This allows a principled account of deictic meaning (e.g. I refers to the speaker, you to the addressee and now to the time of speaking), as shown in Figure 1 and Table 1. Perhaps even more importantly, it is possible to incorporate sociolinguistic information into the grammar, by indicating the kind of person who is a typical speaker or addressee, or the typical situation of use.

Treating utterances as part of the grammar has two further effects which are important for the psycholinguistics of processing and of acquisition. As far as processing is concerned, the main point is that WG accommodates deviant input because the link between tokens and types is guided by the rather liberal `Best Fit Principle' (EWG, 45ff): assume that the current token isa the type that provides the best fit with everything that is known. The default inheritance process which this triggers allows known characteristics of the token to override those of the type; for example, a misspelled word such as mispelled can *isa* its type, just like any other exception, though it will also be shown as a deviant example. There is no need for the analysis to crash because of an error. (Of course a WG grammar is not in itself a model of either production or perception, but simply provides a network of knowledge which the processor can exploit.) Turning to learning, the similarity between tokens and types means that learning can consist of nothing but the permanent storage of tokens minus their utterance-specific content.

These remarks about utterances are summarised in Figure 7, which speculates about my mental representation for the (written) `utterance' Yous mispelled it. According to this diagram, the grammar supplies two kinds of utterance-based information about w1

- that its referent is a set whose members include its addressee,
- that its speaker is a `northerner' (which may be inaccurate factually, but is roughly what I believe to be the case).

It also shows that w2 is a deviant token of the type `MISSPELL: past'. (The horizontal line below 'parts' is short-hand for a series of lines connecting the individual letters directly to the morpheme, each with a distinct part name: part 1, part 2 and so on.)

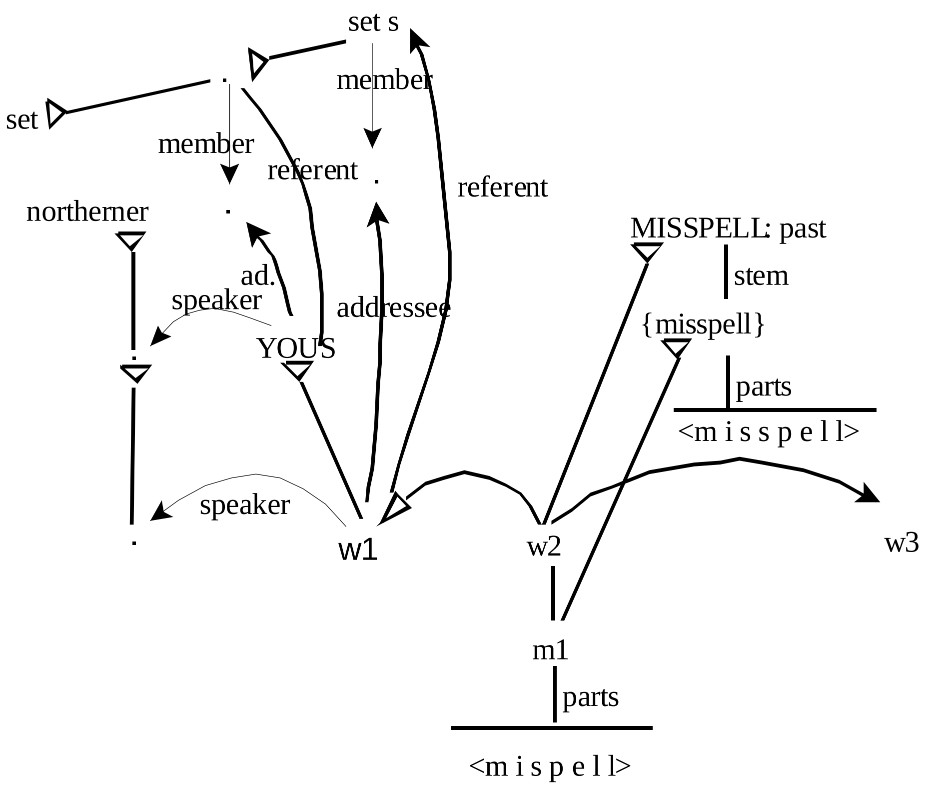


Figure 7

7 Morphology

As explained earlier, the central role of the word automatically means that the syntax is `morphology-free'. Consequently it would be fundamentally against the spirit of WG to follow transformational analyses in taking Jo snores as Jo 'tense' snore. A morpheme for tense is not a word in any sense, so it cannot be a syntactic node. The internal structure of words is handled almost entirely by morphology. (The exception is the pattern found in clitics, which we return to at the end of this section.)

The WG theory of inflectional morphology has developed considerably in the last few years (Creider and Hudson 1998, Hudson 2000a) and is still evolving. At the time of writing (mid 2002) I distinguish sharply between words, which are abstract, and **forms**, which are their concrete (visible or audible) shapes; so I now accept the distinction between syntactic words and phonological words (Rosta 1997) in all but terminology. The logic behind this distinction is simple: if two words can share the same form, the form must be a unit distinct from both. For example, we must recognise a morpheme {bear} which is distinct from both the noun and the verb that share it ($BEAR_{noun}$ and $BEAR_{verb}$). This means that a word can never be directly related to phonemes and letters, in contrast with the EWG account where this was possible (e.g. p. 90: 'whole of THEM = <them>). Instead, words are mapped to forms, and forms to phonemes and letters. A form is the **'shape'** of a word, and a phoneme or letter is a **'part'** of a form. In Figure 7, for example, the verb MISSPELL has the form {misspell} as its stem (a kind of shape), and the parts of {misspell} are <misspell>.

In traditional terms, syntax, form and phonology define different 'levels of language'. As in traditional structuralism, their basic units are distinct: words, morphemes and phoneme-type segments; and as in the European tradition, morphemes combine to define larger units of form which are still distinct from words. For example, {misspell} is clearly not a single morpheme, but it exists as a unit of form which might be written {mis+spell} - two morphemes combining to make a complex form - and similarly for {mis+spell+ed}, the shape of the past tense of this verb. Notice that in this analysis {...} indicates forms, not morphemes; morpheme boundaries are shown by '+'.

Where does morphology, as a part of the grammar, fit in? **Inflectional morphology** is responsible for any differences between a word's stem - the shape of its lexeme - and its whole - the complete shape. For example, the stem of *misspelled* is {misspell}, so inflectional morphology explains the extra suffix. **Derivational morphology**, on the other hand, explains the relations between the stems of distinct lexemes - in this case, between the lexemes SPELL and MISSPELL, whereby the stem of one is contained in the stem of the other. The grammar therefore contains the following 'facts':

- the stem of SPELL is {spell}
- the stem of MISSPELL is {mis+spell}
- the 'mis-verb' of a verb has a stem which contains {mis} + the stem of this verb
- the whole of MISSPELL: past is {mis+spell+ed}
- the past tense of a verb has a whole which contains its stem + {ed}.

In more complex cases (which we cannot consider here) the morphological rules can handle vowel alternations and other departures from simple combination of morphemes.

A small sample of a network for inflectional morphology is shown in Figure 8. This diagram shows the default identity of whole and stem, and the default rule for plural nouns: their shape consists of their stem followed by s. No plural is stored for regular nouns like DUCK, but for GOOSE the irregularity is stored. According to the analysis shown here, geese is doubly irregular, having no suffix and having an irregular stem whose vowel positions (labelled here simply `1' and `2') are filled by (examples of) <e> instead of the expected <o>. In spite of the vowel change the stem of geese isa the stem of GOOSE, so it inherits all the other letters, but had it been suppletive a completely new stem would have been supplied.

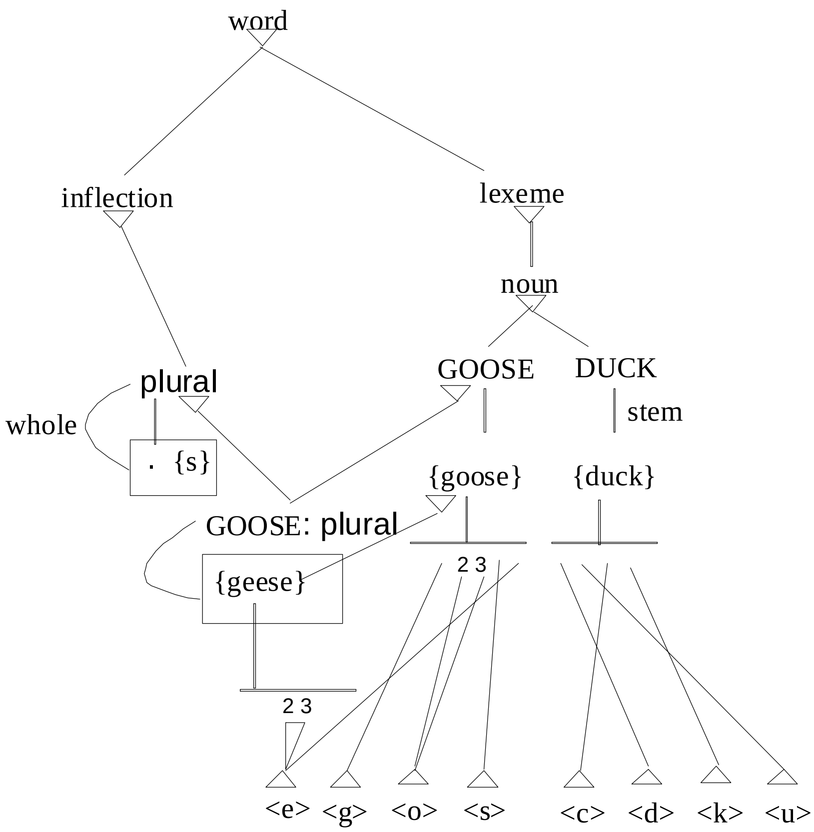


Figure 8

This analysis is very similar to those which can be expressed in terms of `network morphology' (Brown et al 1996), which is also based on multiple default inheritance. One important difference lies in the treatment of syncretism, illustrated by the English verb's past participle and passive participle which are invariably the same. In network morphology the identity is shown by specifying one and cross-referring to it from the other, but this involves an arbitrary choice: which is the `basic' one? In WG morphology, it is possible to introduce further types of 'shape' link such as `en-form'. A word's en-form is one of its shapes alongside its stem and its whole, so the en-form of TAKE is {taken} and that of WALK is {walked}. The en-form is a compromise which allows two word classes (past participle and passive participle) to be mapped regularly onto a range of forms which vary from verb to verb ({ed} by default, {en} for many irregular verbs, no suffix for others, and so on).

As derivational morphology is responsible for relationships between lexemes, it relates one lexeme's stem to that of another. This area of WG is not well developed, but the outlines of a system are clear. It will be based on inter-lexeme relationships such as 'mis-verb' (relating SPELL to MISSPELL) and `nominalisation' (relating it to SPELLING). Derivational morphology is just one kind of **lexical relationship**, in which related lexemes are partly similar in morphology; the grammar must also relate lexemes where morphology is opaque (e.g. DIE - KILL, BROTHER - SISTER). The network approach allows us to integrate all these relationships into a single grammar without worrying about boundaries between traditional sub-disciplines such as derivational morphology and lexical semantics.

I said at the start of this section that clitics are an exception to the generally clear distinction between morphology and syntax. Roughly speaking, a clitic is part word, part affix.

For example, in *He's gone*, the clitic *'s* is a word in terms of syntax, but an affix in terms of morphology. However this is a misleading description if we accept the rigid distinction just presented between words and forms. How can *'s* be both a word and an affix (a kind of morpheme, and therefore a kind of form)? A much more consistent description would be that a clitic is a word whose whole is an affix. They are atypical because typical words have a root; but the exceptionality is just a matter of morphology. In the case of *'s*, I suggest that it *isa* the word BE: present, singular with the one exceptional feature that its whole *isa* the morpheme {s} - exactly the same morpheme as we find in plural nouns, other singular verbs and possessives. As in other uses, {s} needs to be part of a complete word-form which includes a preceding root, so it 'looks for' such a word to its left.

In more complex cases (`special clitics' - Zwicky 1977) the position of the clitic is fixed by the morphology of the host word and conflicts with the demands of syntax, as in the French example (3) where <u>en</u> would follow <u>deux</u> if it were not attached by cliticization to <u>mange</u>, giving a single word-form <u>en mange</u>.

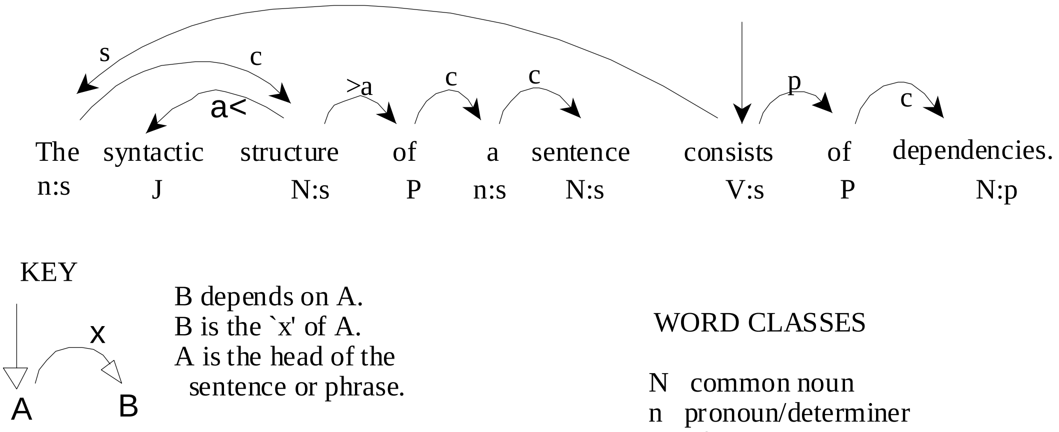
(3) Paul en mange deux.
Paul of-them eats two
`Paul eats two of them.'

Once again we can explain this special behaviour if we analyse *en* as an ordinary word EN whose shape (whole) is the affix {en}. There is a great deal more to be said about clitics, but not here. For more detail see Hudson 2001, Camdzic and Hudson (this volume).

8 Syntax

As in most other theories, syntax is the best developed part of WG, which offers sophisticated explanations for most of the `standard' complexities of syntax such as extraction, raising, control, coordination, gapping and agreement. However the WG view of syntax is particularly controversial because of its rejection of phrase structure. WG belongs to the family of `dependency-based' theories, in which syntactic structure consists of dependencies between pairs of single words. As we shall see below, WG also recognizes `word-strings', but even these are not the same as conventional phrases.

A syntactic dependency is a relationship between two words that are connected by a syntactic rule. Every syntactic rule (except for those involved in coordination) is `carried' by a dependency, and every dependency carries at least one rule that applies to both the dependent and its `parent' (the word on which it depends). These word-word dependencies form chains which link every word ultimately to the word which is the head of the phrase or sentence; consequently the individual links are asymmetrical, with one word depending on the other for its link to the rest of the sentence. Of course in some cases the direction of dependency is controversial; in particular, published WG analyses of noun phrases have taken the determiner as head of the phrase, though this analysis has been disputed and may turn out to be wrong (Van Langendonck 1994). The example in Figure 9 illustrates all these characteristics of WG syntax.



DEPENDENCY TYPES

s subject
c complement
p prepositional
a<, >a pre/post-adjunct

WORD CLASSES

N common noun
n pronoun/determiner
V (full) verb
J adjective
P preposition

:s singular/s-form
:p plural/present

Figure 9

A dependency analysis has many advantages over one based on phrase structure. For example, it is easy to relate a verb to a lexically selected preposition if they are directly connected by a dependency, as in the pair consists of in Figure 9; but it is much less easy (and natural) to do so if the preposition is part of a prepositional phrase. Such lexical interdependencies are commonplace in language, so dependency analysis is particularly well suited to descriptions which focus on 'constructions' - idiosyncratic patterns not covered by the most general rules (Holmes and Hudson 2003). A surface dependency analysis (explained below) can always be translated into a phrase structure by building a phrase for each word consisting of that word plus the phrases of all the words that depend on it (e.g. a sentence; of a sentence; and so on); but dependency analysis is much more restrictive than phrase-structure analysis because of its total flatness. Because one word can head only one phrase it is impossible to build a dependency analysis which emulates a VP node or `unary branching'. This restrictiveness is welcome, because it seems that such analyses are never needed.

In contrast, the extra richness of dependency analysis lies partly in the labelled dependency links, and partly in the possibility of multiple dependencies. In a flat structure, in contrast with phrase structure, it is impossible to distinguish co-dependencies (e.g. a verb's subject and object) by configuration, so labels are the only way to distinguish them. There is clearly a theoretical trade-off between phrase structure and labelled functions: the more information is given in one, the less needs to be given in the other. The general theory of WG is certainly compatible with phrase structure - after all, we undoubtedly use part-whole structures in other areas of cognition, and they play an important role in morphology - but it strongly favours dependency analysis because labelled links are ubiquitous in the cognitive network, both in semantics, and elsewhere. If knowledge is generally organised in terms of labelled links, why not also in syntax? But if we do use labelled links (dependencies) in syntax, phrase structure is redundant.

Syntactic structures can be much more complex than the example in Figure 9. We shall briefly consider just two kinds of complication: structure-sharing, coordination and unreal words. **Structure-sharing** is found when one word depends on more than one other word - i.e. when it is `shared' as a dependent. The notion is familiar from modern phrase-structure analyses, especially HPSG (Pollard and Sag 1994, 19), where it is described as `the central explanatory mechanism', and it is the main device in WG which allows phrases to be discontinuous. (In recognising structure-sharing, WG departs from the European tradition of dependency analysis which generally allows only strictly `projective', continuous structures such as Figure 9.) Figure 10 illustrates two kinds of structure-sharing - in raising (you shared by have and been) and in extraction (what shared by have, been, looking and at). The label `x<' means `extractee', and `r' means `sharer' (otherwise known as `xcomp' or `incomplement').

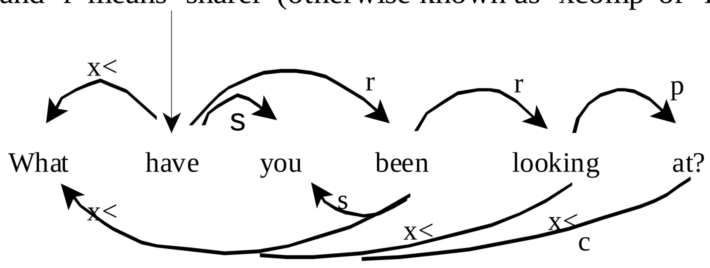


Figure 10

This diagram also illustrates the notion `surface structure' mentioned above. Each dependency is licensed by the grammar network, but when the result is structure-sharing, just one of these dependencies is drawn above the words; the totality of dependencies drawn in this way constitutes the sentence's surface structure. In principle any of the competing dependencies could be chosen, but in general only one choice is compatible with the `geometry' of a well-formed surface structure, which must be free of `tangling' (crossing dependencies - i.e. discontinuous phrases) and `dangling' (unintegrated words). There are no such constraints on the non-surface dependencies. (For extensive discussion of how this kind of analysis can be built into a parsing algorithm, see Hudson 2000c; for a comparison with phrase-structure analyses of extraction, see Hudson 2003c.)

The other complication is **coordination**. The basis of coordination is that conjuncts must share their `external' dependencies - dependencies (if any) to words outside the coordination. The structure of the coordination itself (in terms of `conjuncts' and `coordinators') is analysed in terms of `word-strings', simple undifferentiated strings of words whose internal organisation is described in terms of ordinary dependencies. A word string need not be a phrase, but can consist of two (or more) mutually independent phrases as in the example of Figure 11, where the coordination and conjuncts are bounded by brackets: {[....] [....]}.

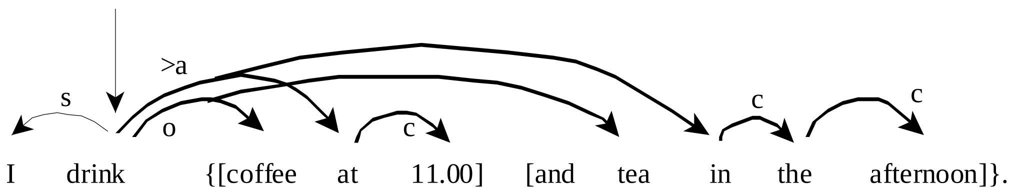


Figure 11

Unreal words are the WG equivalent of 'empty categories' in other theories. Until recently I have rejected such categories for lack of persuasive evidence; for example, my claim has always been that verbs which appeared to have no subject really didn't have any subject at all. So an imperative (*Hurry!*) had no subject, rather than some kind of covert subject. However I am now convinced that this is wrong for at least some languages.

Some languages have case-agreement between subjects and predicatives (WG sharers); for example in Classical Greek the predicative varies in case as illustrated in the examples in (4).

(4) a Kléarkhos phugàs ê:n
Clearchus(nom) exile(nom) was (contrast *phugada*, 'exile(acc)')
'Clearchus was an exile.'

b nomízo: gàr humâ:s emoì eînai kaì patrída kaì phílous
I-think for you(acc) me(dat) to-be and fatherland(acc) and friends(acc)
'for I think you are to me both fatherland and friends' (X. A. 1.3.6)

Notice the nominative case on both subject and predicative noun in (a) contrasting with the accusative in (b). But what if there is no overt subject? In that situation the predicative takes the case that the subject would have had if there had been an overt subject. The subject of an infinitive is always accusative when overt:

(5) emè tatheîn táde
me(acc) to-suffer this
'That I should suffer this!'

As expected, therefore, a predicative in an infinitival clause must also be accusative:

(6) philánthro:pon eînai deî
humane(acc) to-be must
'(one) must be humane'

However the crucial point about examples such as this is that there is no overt subject, so the only way to explain the accusative predicative is to assume a covert one. Similar data can be found in Icelandic and Russian (Hudson 2003d). Creider and Hudson (this volume) discuss the implications for WG theory, and show that 'unreal' words have the same cognitive status as fictions such as Father Christmas.

This discussion of syntax merely sets the scene for many other syntactic topics, all of which now have reasonably well-motivated WG treatments: word order, agreement, features, case-selection, `zero' dependents. The most important point made is probably the claim that the network approach to language and cognition in general leads naturally to dependency analysis rather than to phrase structure in syntax.

9 Semantics

As in any other theory, WG has a compositional semantics in which each word in a sentence contributes some structure that is stored as its meaning. However, these meanings are concepts which, like every other concept, are defined by a network of links to other concepts. This means that there can be no division between `purely linguistic' meaning and `encyclopedic' meaning. For instance the lexemes APPLE and PEAR have distinct senses, the ordinary concepts `apple' and `pear', each linked to its known characteristics in the network of general knowledge. It would be impossible to distinguish them merely by the labels `apple' and `pear' because (as we saw in section 3.2) labels on concepts are just optional mnemonics; the true definition of a concept is provided by its various links to other concepts. The same is true of verb meanings: for example, the sense of EAT is defined by its relationships to other concepts such as `put', `mouth', `chew', `swallow' and `food'. The underlying view of meaning is thus similar to Fillmore's Frame Semantics, in which lexical meanings are defined in relation to conceptual `frames' such as the one for `commercial transaction' which is exploited by the definitions of `buy', `sell' and so on. (See Hudson 2003a for a WG analysis of commercial transaction verbs.)

Like everything else in cognition, WG semantic structures form a network with labelled links like those that are widely used in Artificial Intelligence. As in Jackendoff's Conceptual Semantics (1990), words of all word classes contribute the same kind of semantic structure, which in WG is divided into **`sense'** (general categories) and **`referent'** (the most specific individual or category referred to). The contrast between these two kinds of meaning can be compared with the contrast in morphology (Section 7) between stem and whole: a word's lexeme provides both its stem and its sense, while its inflection provides its whole and its referent. For example, the word dogs is defined by a combination of the lexeme DOG and the inflection `plural', so it is classified as `DOG: plural'. Its lexeme defines the sense, which is `dog', the general concept of a (typical) dog, while its inflection defines the referent as a set with more than one member. As in other theories the semantics cannot identify the particular set or individual which a word refers to on a particular occasion of use, and which we shall call simply `set s'; this must be left to the pragmatics. But the semantics does provide a detailed specification for what that individual referent might be - in this case, a set, each of whose members is a dog. The WG notation for the two kinds of meaning parallels that for the two kinds of word-form: a straight line for the sense and the stem, which are both retrieved directly from the lexicon, and a curved line for the referent and the shape, which both have to be discovered by inference. The symmetry of these relationships can be seen in Figure 12.

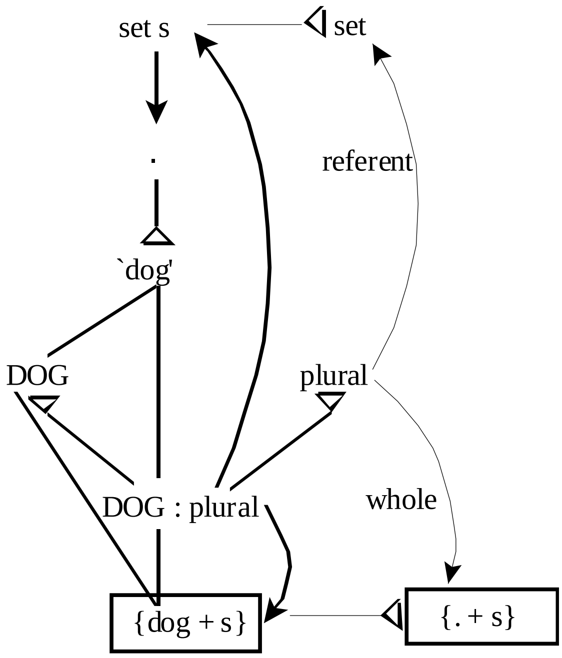


Figure 12

The way in which the meanings of the words in a sentence are combined is guided by the syntax, but the semantic links are provided by the senses themselves. Figure 13 gives the semantic structure for Dogs barked, where the link between the word meanings is provided by `bark', which has an `agent' link (often abbreviated `er' in WG) to its subject's referent. If we call the particular act of barking that this utterance refers to `event-e', the semantic structure must show that the agent of event-e is set-s. As with nouns, verb inflections contribute directly to the definition of the referent, but a past-tense inflection does this by limiting the event's time to some time (`t1') that preceded the moment of speaking (`now'). Figure 13 shows all these relationships, with the two words labelled `w1' and `w2'. For the sake of simplicity the diagram does not show how these word tokens inherit their characteristics from their respective types.

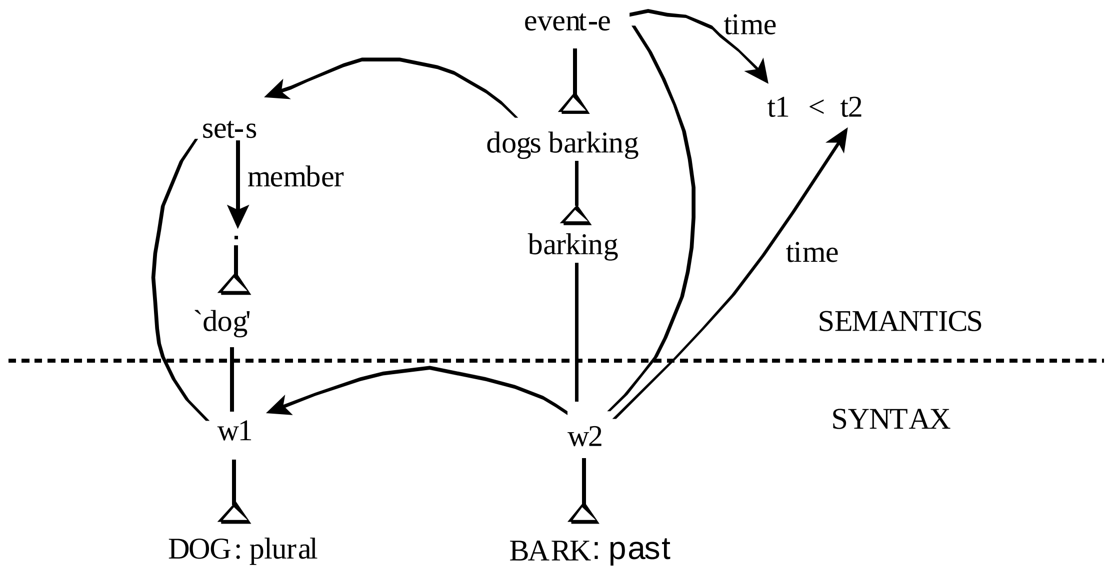


Figure 13

The analysis of Dogs barked illustrates an important characteristic of WG semantic structures. A word's `basic' sense - the one that is inherited from its lexeme - is modified by the word's dependents, which produces a second sense, more specific than the basic sense but more general than the referent. This intermediate sense contains the meaning of the head word plus its dependent, so in effect it is the meaning of that phrase. In contrast with the syntax, therefore, the semantic structure contains a node for each phrase, as well as nodes for the individual words - in short, a phrase structure. Moreover, there are reasons for believing that dependents modify the head word one at a time, each defining a distinct concept, and that the order of combining may correspond roughly to the bracketing found in conventional phrase structure. For example, subjects seem to modify the concepts already defined by objects, rather than the other way round, so Dogs chase cats defines the concepts `chase cats' and `dogs chase cats', but not `dogs chase' - in short, a WG semantic structure contains something like a VP node. This step-wise composition of word meanings is called **`semantic phrasing'**.

This brief account of WG semantics has described some of the basic ideas, but has not been able to illustrate the analyses that these ideas permit. In the WG literature there are extensive discussions of lexical semantics, and some explorations of quantification, definiteness and mood. However it has to be said that the semantics of WG is much less well researched than its syntax.

10 Processing

The main achievements on processing are a theory of parsing and a theory of syntactic difficulty. The most obvious advantage of WG for a parser, compared with transformational theories, is the lack of `invisible' words, but the dependency basis also helps by allowing each incoming word to be integrated with the words already processed, without the need to build (or rebuild) higher syntactic nodes.

A very simple algorithm guides the search for dependencies in a way that guarantees a well-formed surface structure (in the sense defined in section 8): the current word first tries to `capture' the nearest non-dependent word as its dependent, and if successful repeats the operation; then it tries to `submit' as a dependent to the nearest word that is not part of its own phrase (or, if unsuccessful, to the word on which this word depends, and so on recursively up the dependency chain); and finally it checks for coordination. (More details can be found in Hudson 200c.) The algorithm is illustrated in the following sequence of `snapshots' in the parsing of <u>Short sentences make good examples</u>, where the last word illustrates the algorithm best. The arrows indicate syntactic dependencies without the usual labels; and it is to be understood that the semantic structure is being built simultaneously, word by word. The structure after `:-' is the output of the parser at that point.

(1) a w1 = short. No progress:- w1.
b w2 = sentences. Capture:- w1 ← w2.
c w3 = make. Capture:- w1 ← w2 ← w3.
d w4 = good. No progress:- w1 ← w2 ← w3, w4 ←.
e w5 = examples. Capture:- w4 ← w5.
f Submit:- w1 ← w2 ← w3 → (w4 ←) w5.

The familiar complexities of syntax are mostly produced by discontinuous patterns. As explained in section 8, the discontinuous phrases are shown by dependencies which are drawn beneath the words, leaving a straightforward surface structure. For example, subject-raising in <u>He has been working</u> is shown by non-surface subject links from both <u>been</u> and <u>working</u> to <u>he</u>. Once the surface structure is in place, these extra dependencies can be inferred more or less mechanically (bar ambiguities), with very little extra cost to the parser.

The theory of syntactic complexity (Hudson 1996b) builds on this incremental parsing model. The aim of the parser is to link each word as a dependent to some other word, and this link can most easily be established while both words are still active in working memory. Once a word has become inactive it can be reconstructed (on the basis of the meaning that it contributed), but this is costly. The consequence is that short links are always preferred to long ones. This gives a very simple basis for calculating the processing load for a sentence (or even for a whole text): the mean `dependency distance' (calculated as the number of other words between a word and the word on which it depends). Following research by Gibson (1997) the measure could be made more sophisticated by weighting intervening words, but even the simple measure described here gives plausible results when applied to sample texts (Hiranuma 2001). It is also supported by a very robust statistic about English texts: that dependency links tend to be very short. (Typically 70% of words are adjacent to the word on which they depend, with 10% variation in either direction according to the text's difficulty.)

References

Aitchison, Jean (1987): *Words in the Mind. An Introduction to the Mental Lexicon*. Oxford.
Altman, Gerry (1997): *The Ascent of Babel. An Exploration of Language, Mind and Understanding*. Oxford.
Anderson, John (1971): Dependency and grammatical functions. *Foundations of Language* 7, 30-7.
Bates, Elizabeth; Bates,Elizabeth; Elman,Jeffrey; Johnson,Mark; Karmiloff-Smith,Annette; Parisi,Domenico; Plunkett,Kim (1998). On innateness. In William Bechtel and George Graham, A Companion to Cognitive Science. Oxford: Blackwell.
Bennett, David (1994): Stratificational Grammar. In: Asher, Ronald (ed.) *Encyclopedia of Language and Linguistics*. Oxford, 4351-6.
Bresnan, Joan (1978): A realistic transformational grammar. In: Halle, Morris; Bresnan, Joan and Miller, George (eds.) *Linguistic Theory and Psychological Reality*. Cambridge, MA, 1-59.
Brown, Dunstan; Corbett, Greville; Fraser, Norman; Hippisley, Andrew; and Timberlake, Alan (1996): Russian noun stress and network morphology. *Linguistics* 34, 53-107.
Butler, Christopher (1985): *Systemic Linguistics: Theory and Application*. London.
Camdzic, Amela and Hudson, Richard (2002): Clitics in Serbo-Croat-Bosnian. UCL Working Papers in Linguistics 14.
Chekili, Ferid (1982): *The Morphology of the Arabic Dialect of Tunis*. London.
Chomsky, Noam (1957): *Syntactic Structures*. The Hague.
Chomsky, Noam (1965): *Aspects of the Theory of Syntax*. Cambridge, MA.
Chomsky, Noam (1970): Remarks on nominalization. In: Jacobs, Rodney and Rosenbaum, Peter (eds.) *Readings in Transformational Grammar*. London, 184-221.
Creider, Chet (1999) Mixed Categories in Word Grammar: Swahili infinitival nouns. *Linguistica Atlantica* 21, 53-68.
Creider, Chet and Hudson, Richard (1999) Inflectional morphology in Word Grammar. *Lingua* 107, 163-187.
Creider, Chet and Hudson, Richard (2003) Case Agreement in Ancient Greek: implications for a theory of covert elements. This volume.
Evans, Roger and Gazdar, Gerald (1996): DATR: A language for lexical knowledge representation. *Computational Linguistics* 22, 167-216.
Fillmore, Charles (1975): An alternative to checklist theories of meaning. *Proceedings of the Berkeley Linguistics Society* 1, 123-31.
Fillmore, Charles (1976): Frame semantics and the nature of language. *Annals of the New York Academy of Sciences* 280, 20-32.
Fraser, Norman (1985): *A Word Grammar Parser*. London.
Fraser, Norman (1988): Parsing and dependency grammar. *UCL Working Papers in Linguistics* 1, ??
Fraser, Norman (1993): *Dependency Parsing*. London.
Fraser, Norman (1994): Dependency Grammar. In: Asher, Ronald (ed.) *Encyclopedia of Language and Linguistics*. Oxford, 860-4.
Fraser, Norman and Hudson, Richard (1992): Inheritance in Word Grammar. *Computational Linguistics* 18, 133-58.
Gibson, Edward (1997): Linguistic complexity: locality of syntactic dependencies. unpublished.

Gisborne, Nikolas (1993): Nominalisations of perception verbs. *UCL Working Papers in Linguistics* 5, 23-44.
Gisborne, Nikolas (1996): *English Perception Verbs*. London.
Gisborne, Nikolas (2000):The complementation of verbs of appearance by adverbs. In Bermudez-Otero,Ricardo; Denison,David; Hogg,Richard; McCully,C. (eds.) *Generative Theory and Corpus Studies. A dialogue from 10 ICEHL*. Berlin: Mouton de Gruyter, 53-75.
Gisborne, Nikolas (2001):The stative/dynamic contrast and argument linking. *Language Sciences* 23, 603-637.
Goldberg, Adele (1995): *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago.
Gorayska, Barbara (1985): *The Semantics and Pragmatics of English and Polish with Reference to Aspect*. London.
Halliday, Michael (1961): Categories of the theory of grammar. *Word* 17, 241-92.
Halliday, Michael (1967-8): Notes on transitivity and theme in English. *Journal of Linguistics* 3, 37-82, 199-244; 4, 179-216.
Halliday, Michael (1985): *An Introduction to Functional Grammar*. London.
Hiranuma, So (1999): Syntactic Difficulty in English and Japanese: a Textual Study. *UCL Working Papers in Linguistics* 11.
Hiranuma, So (2001): *The syntactic difficulty of Japanese sentences*. UCL PhD.
Holmes, Jasper and Hudson, Richard (2003): Constructions in Word Grammar. In Östman,Jan-Ola; Fried,Mirjam (eds.) *Construction Grammar(s): Cognitive dimensions.*
Hudson, Richard (1964): *A Grammatical Analysis of Beja*. London
Hudson, Richard (1970): *English Complex Sentences. An Introduction to Systemic Grammar*. Amsterdam.
Hudson, Richard (1976): *Arguments for a Non-transformational Grammar*. Chicago.
Hudson, Richard (1980a): *Sociolinguistics*. Cambridge.
Hudson, Richard (1980b): Constituency and dependency. *Linguistics* 18, 179-98.
Hudson, Richard (1981): Panlexicalism. *Journal of Literary Semantics* 10, 67-78.
Hudson, Richard (1984): *Word Grammar*. Oxford.
Hudson, Richard (1989): Towards a computer-testable Word Grammar of English. *UCL Working Papers in Linguistics* 1, 321-39.
Hudson, Richard (1990): *English Word Grammar*. Oxford.
Hudson, Richard (1992): Raising in syntax, semantics and cognition. In: Roca, Iggy (ed.) *Thematic Structure: Its Role in Grammar*. The Hague, 175-98.
Hudson, Richard (1993): Do we have heads in our minds? In: Corbett, Greville; McGlashen, Scott and Fraser, Norman (eds.) *Heads in Grammatical Theory*. Cambridge, 266-91.
Hudson, Richard (1995): *Word Meaning*. London.
Hudson, Richard (1996a): *Sociolinguistics (2nd edition)*. Cambridge.
Hudson, Richard (1996b): The difficulty of (so-called) self-embedded structures. *UCL Working Papers in Linguistics* 8, 283-314.
Hudson, Richard (1997a): The rise of auxiliary DO: verb non-raising or category-strengthening? *Transactions of the Philological Society* 95, 41-72.
Hudson, Richard (1997b): Inherent variability and linguistic theory. *Cognitive Linguistics* 8, 73-108.

Hudson, Richard (1998): *English Grammar*. London.
Hudson, Richard (2000a): **I amn't*. *Language* 76, 297-323.
Hudson, Richard (2000b): Gerunds and multiple default inheritance. *UCL Working Papers in Linguistics* 12, 303-335.
Hudson, Richard (2000c): Discontinuity. *Traitement Automatique des Langues.* 41, 15-56.
Hudson, Richard (2001): Citics in Word Grammar. *UCL Working Papers in Linguistics* 13, 293-294.
Hudson, Richard (2003a): Buying and selling in Word Grammar. In Andor,József; Pelyvás,Peter (eds.) Empirical, Cognitive-Based Studies In The Semantics-Pragmatics Interface. Oxford: Elsevier.
Hudson, Richard (2003b) Mismatches in default inheritance. In Francis,Elaine; Michaelis,Laura (eds.) *Linguistic Mismatch: Scope and Theory*, Stanford: CSLI.
Hudson, Richard (2003c) Trouble on the left periphery. *Lingua*.
Hudson, Richard (2003d) Case-agreement, PRO and structure-sharing. *Research in Language* 1.
Hudson, Richard and Holmes, Jasper (2003) Re-cycling in the Encyclopedia. In Bert Peeters (ed.) The Lexicon/Encyclopedia Interface. Oxford: Elsevier.
Jackendoff, Ray (1990): *Semantic Structures*. Cambridge, MA.
Kaplan, Ron and Bresnan, Joan (1982): Lexical-functional Grammar: a formal system for grammatical representation. In: Bresnan, Joan (ed.) *The Mental Representation of Grammatical Relations*. Cambridge, MA., 173-281.
Kreps, Christian (1997): *Extraction, Movement and Dependency Theory*. London.
Lakoff, George (1987): *Women, Fire and Dangerous Things. What Categories Reveal about the Mind*. Chicago.
Lamb, Sidney (1966): *An Outline of Stratificational Grammar*. Washington, DC.
Lamb, Sidney (1998): *Pathways of the Brain. The neurocognitive basis of language.* Amsterdam: Benjamins.
Langacker, Ronald (1987): *Foundations of Cognitive Grammar I. Theoretical Prerequisites.* Stanford.
Langacker, Ronald (1990): *Concept, Image and Symbol. The Cognitive Basis of Grammar*. Berlin.
Luger, George and Stubblefield, William (1993): *Artificial Intelligence. Structures and Strategies for Complex Problem Solving*. Redwood City.
Lyons, John (1977): *Semantics*. Cambridge.
McCawley, James (1968): Concerning the base component of a transformational grammar. *Foundations of Language* 4, 243-69.
Pinker, Steven (1994): *The Language Instinct*. Harmondsworth.
Pinker, Steven and Prince, Alan (1988): On language and connectionism: Analysis of a Parallel Distributed Processing model of language acquisition. *Cognition* 28, 73-193.
Pollard, Carl and Sag, Ivan (1994): *Head-driven Phrase Structure Grammar*. Chicago.
Reisberg, Daniel (1997): *Cognition. Exploring the Science of the Mind*. New York.
Robinson, Jane (1970): Dependency structures and transformational rules. *Language* 46, 259-85.
Rosta, Andrew (1994): Dependency and grammatical relations. *UCL Working Papers in Linguistics* 6, 219-58.
Rosta, Andrew (1996): S-dependency. *UCL Working Papers in Linguistics* 8, 387-421.

Rosta, Andrew (1997): *English Syntax and Word Grammar Theory*. London.
Shaumyan, Olga (1995): *Parsing English with Word Grammar*. London.
Sugayama, Kensei (1991): More on unaccusative Sino-Japanese complex predicates in Japanese. *UCL Working Papers in Linguistics* 3, 397-415.
Sugayama, Kensei (1992): A word-grammatic account of complements and adjuncts in Japanese (interim report). *Kobe City University Journal* 43, 89-99.
Sugayama, Kensei (1993): A word-grammatic account of complements and adjuncts in Japanese. *Proceedings of the 15th International Congress of Linguistics, Vol 2*. Université Laval, 373-6.
Sugayama, Kensei (1996): Semantic structure of `eat' and its Japanese equivalent `taberu': a Word-Grammatic account. *Translation and Meaning* 4, 193-202.
Taylor, John (1989): *Linguistic Categorisation: An Essay in Cognitive Linguistics*. Oxford.
Tesnière, Lucien (1959) *Éléments de Syntaxe Structurale*. Paris.
Touretsky, David (1986): *The Mathematics of Inheritance Systems*. Los Altos.
Tzanidaki, Dimitra (1995): Greek word order: towards a new approach. *UCL Working Papers in Linguistics* 7, 247-77.
Tzanidaki, Dimitra (1996a): Configurationality and Greek clause structure. *UCL Working Papers in Linguistics* 8, 449-84.
Tzanidaki, Dimitra (1996b): *The Syntax and Pragmatics of Subject and Object Position in Modern Greek*. London.
van Langendonck, Willy (1987): Word Grammar and child grammar. *Belgian Journal of Linguistics* 2, 109-32.
van Langendonck, Willy (1994): Determiners as heads? *Cognitive Linguistics* 5, 243-59.
Volino, Max (1990): *Word Grammar, Unification and the Syntax of Italian Clitics*. Edinburgh.
Zwicky, Arnold (1977): *On Clitics*. Bloomington.
Zwicky, Arnold (1992): Some choices in the theory of morphology. In: Levine, Robert (ed.) *Formal Grammar: Theory and Implementation*. Oxford, 327-71.