19. Word Grammar[i]

Richard Hudson

1. Language as a conceptual network

Word Grammar is a theory of language which touches on almost all aspects of synchronic linguistics and unifies them all through a single very general claim (Hudson 1984: 1):

The Network Postulate:

Language is a conceptual network.

This claim is hardly contentious in Cognitive Linguistics, where it is often taken for granted that language as a whole is a network in contrast with the more traditional view of language as a grammar plus a dictionary - a list of rules or principles and a list of lexical items. However, it is particularly central to Word Grammar, in which each of the main traditional areas of language is a sub-network within the total network of language.

Most obviously, 'the lexicon' is a network of:

- forms
- meanings
- lexemes

(The scare-quotes round 'the lexicon' anticipate section 8, which argues that the lexicon is not an identifiable part of the total language.) This is a network rather than a simple list because the elements among the parts are in a many:many relation. There are lexemes which have more than one meaning (polysemy); there are meanings which are shared by more than one lexeme (synonymy); there are lexemes which have more than one form (inherent variability).

All these mismatches can be illustrated in a single Word: ONE.

- polysemy or homonymy: It means either '1' (contrasting with '2') or 'people' (as in *One shouldn't reveal one's feelings*.).
- synonymy: In the second of these meanings it is synonymous with *you* - which in turn is polysemous.
- inherent variability: regardless of meaning, it has two pronunciations, which in England are /wʌn/ (in the South) and /wɒn/ (in the North). These two pronunciations compete in the speech of those (like me) who have migrated southards. Each of these pronunciations is also available for another Word: *won* or *wan*.

These relationships are most easily shown as a network such as figure 19.1 where no one grouping takes priority as the basis for organising the information.

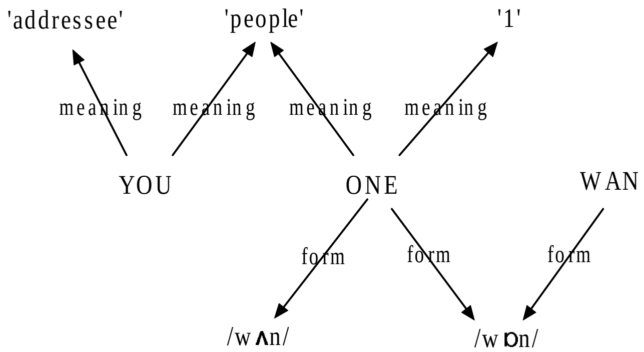


Figure 19.1: A network illustrating polysemy, homonymy, synonymy and variability

Even when applied to the lexicon the network postulate is controversial in comparison with the surprisingly widespread view that the lexicon is organised just like a conventional dictionary (but without the alphabetic order). In this view, the lexicon is a list of lexical items (or lexical entries) each of which combines a single meaning, a word class and a single form (e.g. Jackendoff 1997: 109; Radford 1997: 514). The trouble with this view is that it creates a host of pseudo-questions about the boundaries between the supposed lexical items - e.g. do the two meanings of *one* belong to the same lexical item or to different items? and what about the two pronunciations? It is never defended explicitly against the network view, and probably indicates a lack of interest in these questions rather than a denial of the network view. In contrast, the literature on psycholinguistics commonly presents evidence for the network view, which is now taken as uncontroversial (Aitchison 1997).

At the other end of the spectrum of views, Word Grammar claims that **all** linguistic knowledge has the same basic network architecture. The later sections of this article will show how this claim applies to other areas of language, but we must first consider what it means. What is a network (in the Word Grammar sense), and what does it contrast with?

2. Networks as notation

At one level, the Network Postulate may be thought of simply as a claim about the notation for displaying linguistic data. Seen at that level, a network is a graph consisting of a set of nodes and a set of lines. According to Word Grammar, the formal properties of such a graph are as follows:

- Each node must be connected by lines to at least two other nodes (otherwise it would be a mere dot, rather than a node where two lines meet).
- There are two kinds of line (either of which may be either straight or curved):
  - 'isa' lines (showing class-member relations), with a small triangle at one end,
  - arrows.
- An isa line has either a node at each end or an arrow at each end.
- An arrow points from either a node or an arrow to a node (which may be the same as the source node).
- The nodes are all labelled as:
  - constants (shown as a mnemonic name) or
  - variables (shown either as a number between 0 and 1, or simply as an unlabelled dot).
- The lines are all labelled as constants.

As we shall see in section 7, the individual labels are in fact redundant, but the distinction between variables and constants is (probably) not.

These formal characteristics of a Word Grammar network are illustrated abstractly in figure 19.2.

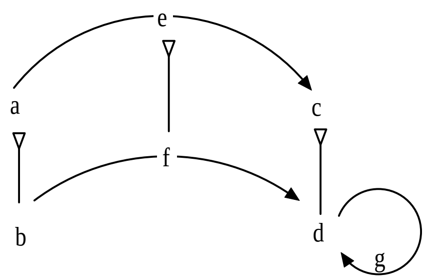


Figure 19.2: An abstract illustration of the notation of Word Grammar.

The notation has an unambiguous semantics:

- A triangle-based line shows an 'isa' (classification) relation in which the triangle rests on the super-category:
  - b isa a
  - d isa c
  - f isa e
- An arrow points from one node to the node that has the named relation to it - in other words, it is a function from the first node to the second:
  - the e of a is c
  - the f of b is d
  - the g of d is d

Word Grammar claims that this notation applies throughout language, from phonology through morphology and syntax to semantics and sociolinguistics. The claim that a single notation suffices for all levels of language is itself a significant part of Word Grammar theory, because it is implicitly denied by the plethora of different notations which are currently thought necessary for analysing different kinds of linguistic structures. The following list is not exhaustive:

- trees
- stemmas
- attribute-value matrices
- directed acyclic graphs
- multi-tiered phonological and morphological structures
- linear strings of phonemes
- bracketed strings (of words or logical symbols), with or without labelling.

3. Networks as theory

However the Network Postulate is not merely a matter of notation. It also implies a theory of language structure with a number of specific sub-theories, which will be discussed briefly below.

- conceptual distance and activation
- entrenchment
- openness
- declarativeness

In general the Word Grammar position on networks is typical of Cognitive Linguistics (e.g. Barlow, Kemmer 2000; Goldberg 1995; Langacker 1998; Langacker 2000), though the notation and some details differ. However its historical roots are much earlier than the start of Cognitive Linguistics, in Stratificational Grammar (Lamb 1966; Lamb 1998) and Systemic Grammar (Halliday 1961; Hudson 1971).

Conceptual distance and activation

It is a commonplace of cognitive psychology that knowledge is a network, and that the network supports spreading activation in which activation of one node 'spills over' to neighbouring nodes (Reisberg 1997: 256-303). In short, the network allows the 'conceptual distance' between nodes to be represented, so that some nodes are nearer to each other than to others; and the relative distances between nodes explain differences in mutual activation. There is a wealth of evidence that words activate ('prime') each other, and some evidence that the same is true for more general grammatical categories - so-called 'structural priming' (Bock, Griffin 2000).
The Network Postulate gives the simplest possible explanation for spreading activation in language: it happens because this is how our brains use networks, and language is a network. In contrast, spreading activation would be hard to explain if language consisted of a list of unrelated lexical items plus a set of rules or principles for combining them.

Moreover, the Network Postulate generates research questions which simply do not arise otherwise; for example, why activation is directional in noun-noun compounds such as *crocodile shoes*, where *crocodile* primes *shoes*, but *shoes* does not prime *crocodile* (Harley 1995: 84; Roelofs 1996). It is not hard to think of possible explanations for this asymmetry in terms of sequential order (earlier primes later) or dependency (dependent primes head) or even the 'fan' effect (the fewer links a node has, the more activation passes through each one). No doubt these alternatives can be distinguished experimentally, but the point is that the question becomes a matter of interest for linguists only if we assume that our theories of language have something to do with spreading activation. This hypothesis has recently led to a great deal of important work by cognitive linguists in areas such as language acquisition (Tomasello 2003) and diachronic change (Bybee 2001).

Entrenchment

Another closely related commonplace of cognitive psychology and psycholinguistics is that the accessibility of stored information varies from concept to concept according to how often we access this particular item of information, giving a 'recency' effect and a 'frequency' effect.

For example, we have variable difficulty in retrieving people's names (and other attributes), in retrieving past tenses of verbs (e.g. the past tense of *thrive* is less accessible than that of *drive*), and so on. These differences cannot be explained in terms of conceptual distance, since the distance (at least in a network analysis) is constant. Nor can they be explained in terms of accessibility of the target concept itself; for example, a name that we cannot recall may turn out to be a very common one. The explanation must lie in the link between the source (e.g. the person) and the target (their name), and specifically in its degree of 'entrenchment'. The term is borrowed from Langacker, who generally uses it to refer to the familiarity or automaticity of a concept rather than of a link (Langacker 2000); it remains to be seen whether this difference is important.

Once again this kind of variation can be explained in terms of a network model, since links may have different degrees of 'entrenchment' reflecting differences of experience - most obviously differences of frequency: the more we use a link, the easier it is to use. In order to show entrenchment, then, we need to be able to treat 'entrenchment' as a property of network links, which presupposes that the analysis includes the links as elements that can carry properties. Network models do include them but others may not.

Openness

A further general characteristic of networks is their lack of natural boundaries, either internal or external. There are clear sub-networks of words which are more or less closely related to one another in terms of single criteria such as meaning, word class, morphology or phonology, but typically the networks defined by one criterion cut across those defined in other ways. Equally, the network of language itself has no clear natural boundaries. This is most obvious where phonology fades into phonetics and where semantics fades into encyclopedic and contextual knowledge: Are the details of allophonic realisation part of language (phonology) or not (phonetics)? And how much of word meaning belongs to language?

The lack of clear boundaries is as expected if the Network Postulate is right, but hard to explain if language consists of a collection of linguistic rules and lexical items. The traditional rules-and-items view is closely related to the scholarly tradition in which each language is described in at least two distinct books - a grammar and a dictionary - and in which general knowledge is assigned to a third kind of book - an encyclopedia. These traditional boundaries are perpetuated in the popular idea of 'modularity' according to which there is a discrete part of the mind, called a module, either for the whole of language or for each of the supposed parts of language (Chomsky 1986; Fodor 1983). This rather crude kind of modularity has always been highly contentious (Garfield 1987), but it is fundamentally incompatible with the Network Postulate. In contrast, the Network Postulate allows and perhaps even encourages a more subtle kind of modularity in which nodes cluster into relatively dense sub-networks, but without absolute boundaries. This is what has been called 'hierarchical modularity' in recent work on the mathematics of networks (Barabási 2003: 236).

Declarative knowledge

A final consequence of the Network Postulate is that knowledge of language is entirely declarative (rather than procedural). This must be so if the relevant knowledge consists of nothing but interconnected nodes; it is simply not possible to formulate a procedure in such terms. A network is like a map which lays out the possible routes, in contrast with a procedure for getting from one place to another. This does not of course mean that language use is irrelevant - far from it. Language use involves activation of the network and even the creation of new nodes and links (i.e. learning). But the Network Postulate distinguishes this activity conceptually from the network to which it is applied.

Of course it is a matter of debate (and ultimately of fact) whether knowledge of language really is entirely declarative. Among those who distinguish rules and lexical items there are many who believe that some or all of the rules are procedures of the form 'If X is true, do Y' (i.e. 'productions'). This is especially true in phonology (e.g. Halle, Bromberger 1989) but has been at least implicit in syntax since Chomsky's first introduction of rewrite rules. If some linguistic knowledge really does turn out to be procedural, the Network Hypothesis will have to be revised or abandoned.

4. 'Isa', default inheritance and prototypes

One particularly important type of link in a Word Grammar network is the ***'isa'*** link, the relationship between a concept and a super-category to which it belongs; for example, the link between the concepts Dog and Animal, or between the word DOG and the word -class Noun. This is the basis for all classification in Word Grammar, regardless of whether the classified concept is a sub-class (e.g. Dog isa Animal) or an individual (e.g. Fido isa Dog), and regardless of whether it is a regular or an exceptional member. All theories in the Cognitive Linguistics tradition recognise classification relations, but the terminology varies - the term 'isa', borrowed from Artificial Intelligence (Reisberg 1997: 280), is only used in Word Grammar - and Cognitive Grammar recognises different relationships for regular and exceptional members (Langacker 2000).

'Isa' relationships are important because of their role in the basic logic of generalisation: *default inheritance* (which is also recognised, with differences of terminology and some details, across Cognitive Linguistics).

Default Inheritance:

Inherit all the characteristics of a super-category unless they are overridden.

Default logic allows generalisations to have exceptions, so in essence if not in name it has been the basic logic of linguistic analysis since the Greek and Sanskrit grammarians. However it is also arguably the logic of ordinary common-sense reasoning, whereby we can recognise a three-legged cat as an exceptional cat rather than a non-cat, or a penguin as an exceptional bird.

In simple terms, if we know that A isa B, and that B has some characteristic C, then we normally assume that A too has C (i.e. A inherits C from B by default). However there is an alternative: if we already know that A has some characteristic which is incompatible with C, this is allowed to 'override' the 'default' characteristic. For example, if we know that A isa Cat, and that Cat (i.e. the typical cat) has four legs, we would normally inherit four-leggedness for A as well; but if we already know that A has only three legs (which is incompatible with having four) we accept this instead of the default number. Similarly, if we know that a typical past-tense verb has the suffix *-ed*, we inherit this pattern for any past-tense verb unless we already know that it does not contain *-ed* (e.g. *took*). Figure 19.3 illustrates both these cases, using the Word Grammar notation explained earlier in which the small triangle indicates an 'isa' relationship. (The examples are of course simplified.) All the links shown with solid lines are stored, but those with dotted lines are inherited.

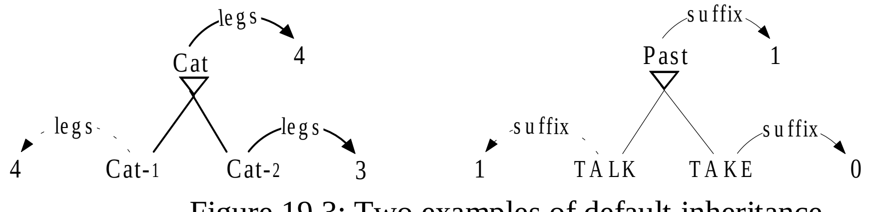


Figure 19.3: Two examples of default inheritance

The default inheritance of Word Grammar allows multiple inheritance - simultaneous inheritance from more than one super-category. For example, Cat isa both Mammal and Pet, so it inherits various bodily characteristics from Mammal and functional characteristics from Pet. In language, multiple inheritance applies most obviously in inflectional morphology; for example, the past tense of TALK isa both TALK and Past, inheriting its stem from TALK and its suffix from Past. This multiple inheritance is unrestricted, so in principle it is possible to inherit conflicting characteristics from two super-categories, leading to a logical impasse. This is proposed as the explanation for the strange gap in English morphology where we expect to find **amn't* (Hudson 2000a).

Although the basic ideas of default inheritance are widely accepted in Cognitive Linguistics, they are not generally invoked in discussions of another leading idea of Cognitive Linguistics, that categories exhibit prototype effects (Barsalou 1992: 162). One distinctive characteristic of a prototype category is that its members have different degrees of typicality (e.g. a penguin is an untypical bird), a variation which is to be expected if we allow default characteristics to be overridden in the case of exceptional examples. The stored characteristics of penguins override some of the default bird characteristics such as flying and being about the size of a sparrow, but these exceptional characteristics do not prevent it from being classified as a bird. The advantage of invoking default inheritance as an explanation of prototype effects is that it removes the need to assume that concepts are themselves fuzzy (Sweetser 1987). Rightly or wrongly, the structure of a Word Grammar network is crystal clear and fully 'digital' (except for degrees of entrenchment and activation).

5. The Best Fit Principle and processing

A further benefit of default inheritance is the possibility of an efficient classification in which the needs of generalisation outweigh those of strict accuracy and reliability. If we know that something is a cat we can inherit a great deal of information about it - e.g. that it enjoys being stroked and hunting small birds - even though some parts of this inherited (inferred) information may turn out to be unreliable. Most of the time, most inherited information is true, and the information flows extremely fast; we sacrifice total reliability for the sake of speed and quantity. The price we pay includes prejudice and the occasional accident.

However there is another cost to be recognised, which is the increased difficulty of processing incoming experiences. What if the bit of experience that we are currently processing turns out to be exceptional? This is allowed by default inheritance, which allows mismatches between tokens and the types to which we assign them; and it is clearly part of our everyday experience. We are often confronted by late buses and sometimes even by three-legged cats, and in language we have to cope with mis-spelt words, foreign pronunciations and poetry.

How, then, do we classify our experiences? The most plausible answer is that we apply the Best Fit Principle (Winograd 1976; Hudson 1984: 20), which favours the classification that gives the best overall 'fit' between the observed characteristics of the experience and some stored category.

The Best Fit Principle:

> Classify any item of experience so as to maximise the amount of relevant inherited information and to minimise the number of exceptions.

This principle allows us to classify a three-legged cat as a cat because all the other observable characteristics match those that we expect from a cat. It is true that we could avoid conflicting features altogether by pitching the classification at a much higher level, say at the level of Thing; although it is an exceptional cat and even an exceptional animal, it is not an exceptional thing; but classifying it merely as a thing would lose the benefits of being able to predict its behaviour - e.g. its reaction to being stroked.

This principle has many attractions, not least its intuitive explanatory power. It also explains another characteristic of categorisation which is part of the theory of prototypes, namely the existence of borderline categories and of categories whose borders shift from context to context. For example, is a particular person a student? It all depends on what kind of contrast we are assuming - between students and graduates, between students and prospective students, between officially registered students and others, and so on. This is as predicted by the Best Fit Principle, because relevance varies with context (Sperber, Wilson 1995).

However this very powerful and attractive theory again has a considerable price. How does it work? Do we really compute all the possible alternative classifications and then select the winner? This cannot possibly be true, because there are so many 'possible alternatives' in any full-sized network of concepts, and yet we classify our experiences almost instantaneously.

Interestingly, the theory of default inheritance also raises a similar problem. If any characteristic may be overridden, how do we know whether or not a particular characteristic actually is overridden in any given case where we might inherit it? Once again the answer seems to involve an exhaustive search of at least a large section of the network.

Both these search problems allow the same plausible solution: *spreading activation*. As explained earlier, we already assume that this is the basis for all processing, so we can assume that at any given moment a small subset of all the nodes in the network are active (or above some threshold of activation). The solution to both the search problems is to assume that the search can be confined to the concepts that are currently active. This solution applies to the Best Fit Principle because all the relevant candidates must be active, so the problem is just to select the active node which provides the most inheritable information - which means, in effect, the most specific one (e.g. Cat rather than Thing). Similarly, the solution also applies to Default Inheritance because any possible overriding node must already be active, so all other nodes in the network may safely be ignored.

6. Classified relations

A Word Grammar network is not a mere associative network which just shows whether or not two nodes are related. Every link in the network is classified. One class of relations is the basic 'isa' relation discussed above, but there are many others - 'wife', 'name', 'nationality', 'meaning', 'subject' and so on. This is normal practice in Cognitive Linguistics, though Word Grammar may be the only theory which regularly uses the arrow notation illustrated in the previous diagrams.

However Word Grammar offers a solution to a general problem that faces network analyses: how to cope with the potential proliferation of relationships (Reisberg 1997: 280). Once we start distinguishing one relationship from another, where do we stop? There are no obvious stopping points between very general relationships such as 'part' and very specific ones such as 'small toe on the left foot'; for example, we are clearly capable of understanding the sentence *He touched the* *small toe on his left foot*, which defines a unique relationship between him and the toe in question, so such specific relationships do in fact seem to be needed in a cognitive network.

The Word Grammar solution is to treat relationships themselves as concepts, and to allow them to be classified and sub-classified just like other concepts. This produces a hierarchy of relationships linked by 'isa' and interpreted by Default Inheritance, as illustrated in figure 19.4. This hierarchy naturally includes the most general relationships such as 'part', but it may extend downwards without limit to include the most specific imaginable relationships, such as that between me and the small toe on my left foot. Since every relationship is a unique example of its super-type, this has the effect of making every relationship into a function - a relationship which has a unique value for any given argument. For example, you and I both have a unique relationship to our left small toe, but these relationships are distinct and are united only in being instances of the same more general relationship.

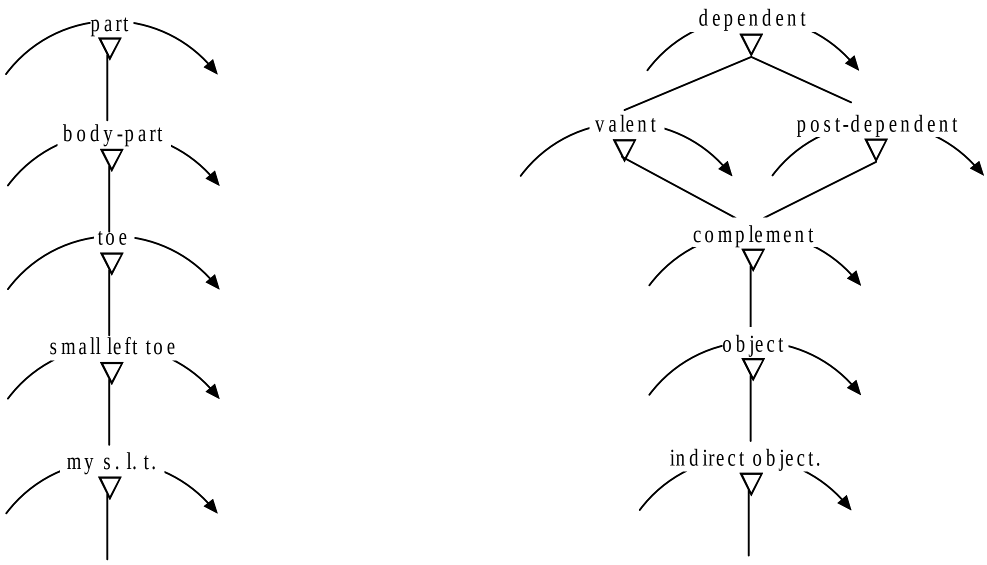


Figure 19.4: Two classification hierarchies of relationships

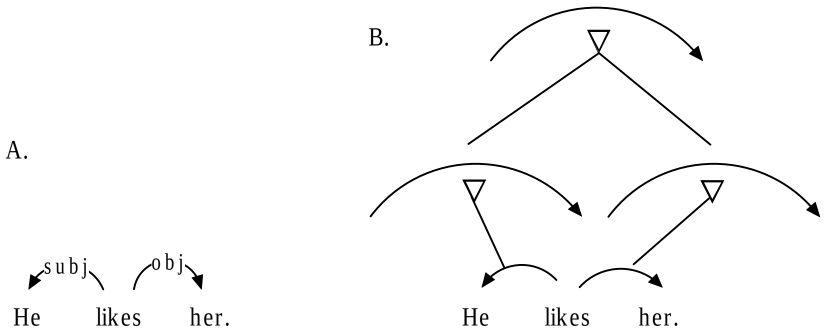
This hierarchical approach to relationships is most obvious in the Word Grammar treatment of grammatical relations - for example, Indirect object isa Object which isa Complement which isa Post-dependent and Valent (i.e. a non-adjunct) which isa Dependent. This classification is also shown in figure 19.4. However similar hierarchies can be found throughout the relationships which are needed for language and (no doubt) elsewhere.

From a formal point of view, this classification of links makes Word Grammar networks very complex compared with most network models because it defines a 'second-order' network of relationships among relationships. Fortunately the second-order relationships are all of the same kind - 'isa' - so they are not likely to lead eventually to a third-order network with a danger of infinite complexity.

7. Labels and uniqueness

This hierarchical analysis of link-types solves another problem. One of the characteristics of a network is that the nodes are defined **only** by their links to other nodes; for instance, the word CAT is the only word that is linked to the concept Cat, to the pronunciation /kat/, to the word -class Noun, and so on. No two nodes have exactly the same links to exactly the same range of other nodes, because if they did they would by definition be the same node. As Lamb points out (Lamb 1966; Lamb 1998: 59), one consequence of this principle is that the labels on the nodes are entirely redundant, in contrast with non-network approaches, in which labels are the only way to show identity. For example, if two rules both apply to the same word -class, this is shown by naming this word -class in both rules; as we all know, the name chosen does not matter, but it is important to use the same name in both rules. In a network, on the other hand, labels only serve as mnemonics to help the analyst, and they could (in principle) all be removed without loss of information.

If we follow the logic of this argument by removing labels from nodes, we face a problem because the labels on *links* appear to carry information which is *not* redundant, because the links are not distinguished in any other way. This leads to a paradoxical situation in which the elements which traditionally are always labelled need not be, but those which (at least in simple associative networks) are traditionally not labelled must be labelled. The hierarchical classification of links resolves this paradox by giving links just the same status as nodes, so that they too can be distinguished by their relationships to other links - i.e. by their place in the overall classification of links. By definition, every distinct link must have a unique set of properties, so a link's properties are always sufficient to distinguish it and labels are redundant. In principle, therefore, we could remove their labels too, converting a labelled diagram such as the simple syntactic structure in figure 19.5A into the unlabelled one in figure 19.5B. (The direction of the arrows in B is arbitrary and does not indicate word order, but the two intermediate arrows in this diagram must carry distinct features, such as word order, to make each one unique.)



Figures 5A and 5B: Distinguishing relationships with and without the use of labels.

8. 'The lexicon' , 'the grammar' and constructions

As in other Cognitive Linguistics theories, there is no distinction in Word Grammar between the lexicon and the grammar. Instead the isa hierarchy of words covers the full range of concepts and facts from the most general facts to the most specific, with no natural break in the hierarchy between 'general' and 'specific'. As we have just seen, the same is true of dependency relationships, where the specific dependencies found in individual sentences are at the bottom of the hierarchy capped by the very general relationship, 'dependent'. There is no basis, therefore, for distinguishing the lexicon from the grammar in terms of levels of generality, because generality varies continuously throughout the hierarchy.

Take again the sentence *He likes her*. One requirement for any grammar is to predict that the verb *likes* needs both a subject and an object, but the rules concerned vary greatly in terms of generality. The subject is needed because *likes* is a verb, whereas the object is needed because it is a form of the lexeme LIKE; in traditional terms, one dependency is explained by 'the grammar' while the other is a 'lexical' fact, so different mechanisms are involved. In Word Grammar, however, the only difference is in the generality of the 'source' concept. Figure 19.6 shows how *likes* inherits its two dependencies from the network.

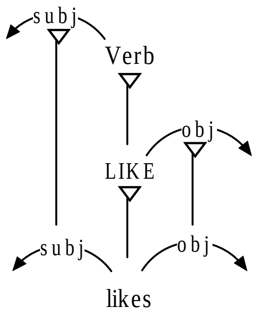


Figure 19.6: Where the subject and object of *likes* come from.

This approach to lexico-grammar solves the problem of what may be called 'special constructions', syntactic patterns which are fully productive but which do not fit any of the 'ordinary' patterns and are tied to specific lexical items (Hudson 1984: 4; Holmes, Hudson 2001forthcoming). For example, the preposition WITH can be used as the root of a sentence provided that it is preceded by a direction expression and followed by a noun phrase.

(1) Down with the government!

(2) Off with his head!

(3) Away with you!

(4) Into the basket with the dirty washing!

This construction is not generated by the rules for normal sentences, but a grammar/lexicon split forces an arbitrary choice between a 'grammatical' and a 'lexical' analysis. In Word Grammar there is no such boundary, and no problem. The sub-network in figure 19.7 provides the basis of an analysis.

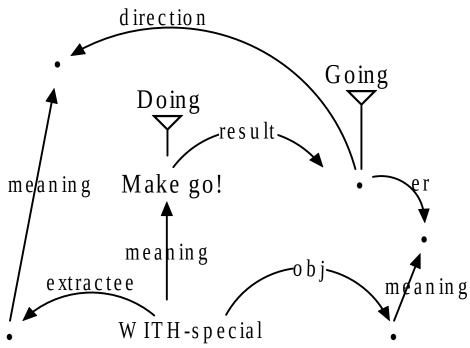


Figure 19.7: A network for the 'X WITH Y' construction

In words, what figure 19.7 says is that WITH-special (this special case of the lexeme WITH) means 'Do something to make Y go to X', where Y is the meaning (referent) of the object noun and X is that of the 'extracted' (front shifted) word . (The relation 'er' in the semantics stands for 'go-er'.) Given the ordinary grammar for noun phrases and directionals, this pattern is sufficient to generate the examples in (1-4), but some parts of the pattern could be omitted on the grounds that they can be inherited from higher nodes which are partly 'grammatical' (e.g. the word classes permitted as object) and partly 'lexical' (e.g. the fact that WITH has an obligatory object).

9. Morphology

The Word Grammar treatment of morphology separates two separate analyses:

- the analysis of word structure in terms of morphemes or phonological patterns;
- the linkage between word structure and lexeme or word class.

For example, the recognition of a suffix in *dogs* is separate from the recognition that *dogs* is plural. The plurality and the suffix are clearly distinct - one is a morpheme, i.e. a word -part, while the other is a word class, and either can exist without the other (as in the plural *geese* or the singular *news*). In this sense, therefore, Word Grammar morphology belongs firmly within the 'word and Paradigm' tradition in which a word's internal structure is distinguished sharply from its morpho-syntactic features (Robins 1959/2001).

The theory of morphology raises a fundamental question about the architecture of language: how many 'levels of analysis' are there? This actually breaks down into two separate questions:

- Is there a 'syntactic' level, at which we recognise words?
- Is there a 'morphological' level, at which we recognise morphemes?

Word Grammar recognises both of these levels (Creider, Hudson 1999), so the relation between semantic and phonological structure is quite indirect: meanings map to words, words to morphemes and morphemes to phonemes (or whatever phonological patterns there are). There is a range of evidence for this view:

- Words and morphemes are classified differently from each other and from the meanings they signal - meanings may be actions, words may be verbs and morphemes may be roots; and morphological 'declension classes' are distinct from morpho-syntactic features.
- Morphological patterns are different from those of syntax; for example, there is no syntactic equivalent of semitic interdigitation (whereby the plural of Arabic *kitaab*, 'book', is *kutub*), nor is there a morphological equivalent of coordination or extraction; many languages have free word order, but none have free morpheme order.
- The units of morphology need not match those of syntax; for example, French syntax recognises the two-word combination *de le* (of the) which corresponds to a single morphological unit *du* (see figure 19.11 below).

This position is quite controversial within linguistics in general, and within Cognitive Linguistics in particular. Cognitive Grammar at least appears to deny the level of syntax, since its symbolic units are merely a pairing of a semantic pole with a phonological pole (Langacker 2000: 5) so they cannot be independently categorised (e.g. in terms of non-semantic word classes). But even if the symbolic units do define a level of syntax, there is certainly no independent level of morphology: "... a basic claim of Cognitive Grammar, namely that morphology and syntax form a continuum (fully describable as assemblies of symbolic structures)." (Langacker 2000:25). In other words, in contrast with the Word and Paradigm model, morphology is merely syntax within the word.

In Word Grammar, then, the word is linked to its phonological realisation only via the morpheme, just as it is linked to the semantic structure only via the single concept that acts as its sense. The pattern for the word *cat* (or more precisely, the lexeme CAT) is shown in figure 19.8. We follow a fairly traditional notation for distinguishing levels: 'Cat' is the concept of the typical cat, 'CAT' is the lexeme, {cat} is the morpheme and /k a t/ are the phonemes.

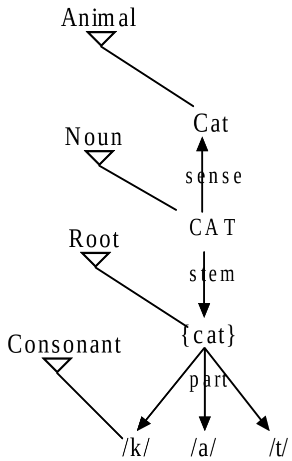


Figure 19.8: The word *cat* analysed on four linguistic levels

Morphologically complex words map onto more than one morpheme at a time, so we need to recognise a complex unit at the level of morphology, the 'word form' (or 'morphological word' - Rosta 1997). For example, the word form {{kat} + {s}} realises the word 'CAT: plural' (the plural of CAT), which isa both CAT and another word type, Plural. These two word-types contribute respectively its stem and its suffix, so in all there are three links between CAT:plural and its morphology:

- the 'stem' link to the stem morpheme, inherited from CAT;
- the 'suffix' link to the suffix morpheme, inherited from Plural;
- the 'word form' link to the entire word form, also inherited from Plural.

These three links are called 'morphological functions' - functions from a word to specific parts of its morphological structure (Hudson 2000a).

Irregular forms can be accommodated easily thanks to default inheritance, as shown in figure 19.9. The top part of this figure (above the dotted line) represents stored information, while the bottom part is information that can be inferred by default inheritance. In words, a plural noun has a word form ('wf' in the diagram) whose first and second parts are its stem and its suffix (which is {s}). The stem of CAT is {cat}, so the whole of CAT: plural consists of {cat}followed by {s}. Exceptionally, the whole of PERSON: plural is stipulated as {people}; by default inheritance, this stipulation overrides the default. As in other areas of knowledge, we probably store some regularly inheritable information such as the plural of some very frequent nouns as well as the unpredictable irregular ones (Bybee 1995).

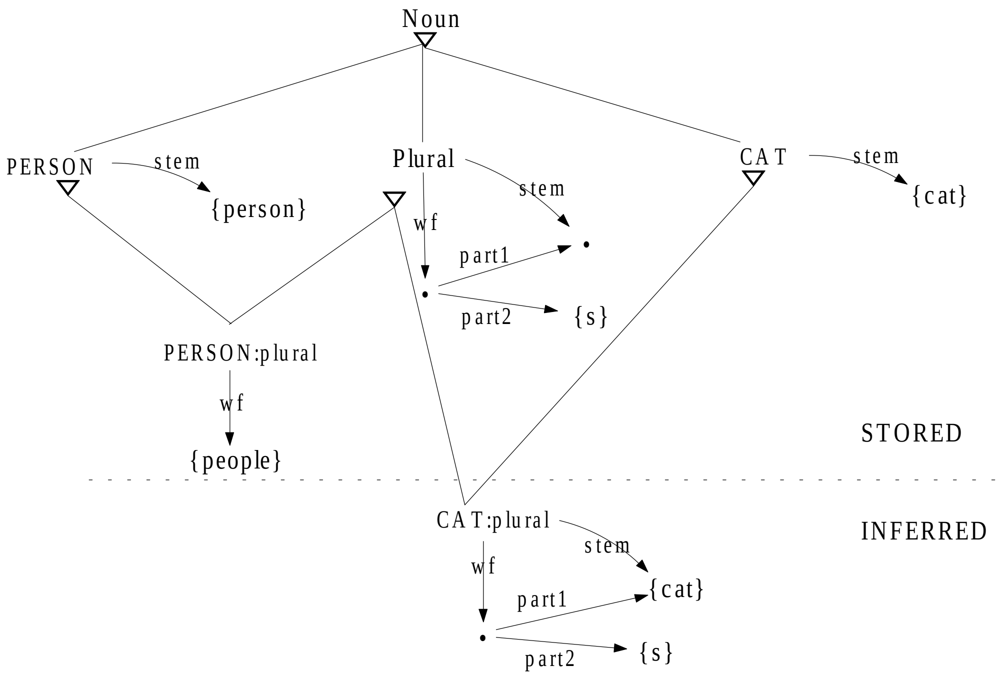


Figure 19.9: Inflectional morphology for English regular and irregular plural nouns

Derivational morphology uses the same combination of morphemes and morphological functions, but in this case the morphology signals a relationship between two distinct lexemes, rather than between a lexeme and an inflectional category. For example, take the agentive pattern in SPEAK - SPEAKER. The relationship between these two lexemes exemplifies a more general relationship between verbs and nouns. The relevant part of the grammar, figure 19.10, shows how lexemes which are related in this way differ in terms of meaning, syntax and morphology. In words, a verb typically has an 'agentive':

- which isa Noun,
- whose stem consists of the verb's stem combined with the {er} suffix,
- whose sense is a person who is the agent ('er') of an event which isa the verb's sense.

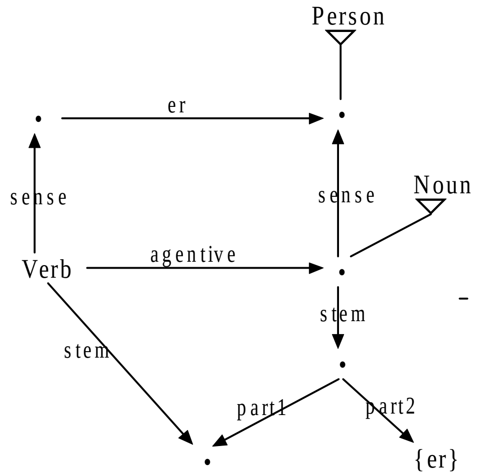


Figure 19.10: Derivational morphology for agentive nouns

One of the benefits mentioned earlier of the distinction between words and their morphological realisation is the possibility of gross mismatches between them, as discussed extensively in Sadock 1991. Figure 19.11 illustrates the familiar case from French of *du*, a single morpheme which realises two words:

- the preposition *de*, 'of',
- the masculine definite article which is written *le*.

For example, alongside *de la fille*, 'of the daughter', we find *du fils*, 'of the son', rather than the expected **de le fils*. One of the challenges of this construction is the interaction between morphology and phonology, since *du* is not used when *le* is reduced to *l'* before a vowel: *de l'oncle*, 'of the man'. The analysis in figure 19.11 meets this challenge by distinguishing the 'full stem' and the 'short stem', and applying the merger with {de} only to the former. Some other rule will prevent the full stem from occurring before vowels, thereby explaining the facts just outlined. The analysis also ensures that *de le* only collapses to *du* when the *le* depends directly on the *de*, as it would in *du fils*.

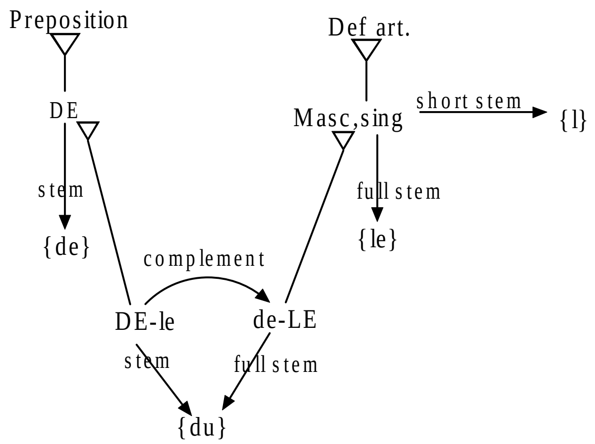


Figure 19.11: Why French *de* le is realised as *du*

10. Syntax

Syntax is the area in which Word Grammar has been applied most thoroughly (Hudson 1984; Hudson 1990; Hudson 1998; Hudson 1999; Hudson 2000c; Hudson 2000b2003a, b), so the following discussion can be quite brief.

The most distinctive characteristic of syntax in Word Grammar is that it assumes dependency structure rather than phrase structure. The syntactic structure of a sentence is a network of nodes, in which there is a node for each word but no nodes for phrases; and the nodes are all connected by syntactic dependencies. For example, in the sentence *Small babies cry,* the noun *babies* depends on the verb *cries*, and the adjective *small* depends on *babies*, but there is no node for the noun phrase which consists of *babies* plus its dependent. It would be easy to add phrase nodes, because they can be read unambiguously off the dependencies, but there is no point in doing so because they are entirely redundant. This way of viewing sentence structure exclusively in terms of word-word dependencies has a long history which goes back through the mediaeval European and Arabic grammarians to classical Greece, but it has recently been overshadowed by the phrase-structure approach (Percival 1976, 1990).

One of the advantages of the dependency approach is that grammatical functions such as 'subject' and 'object' are sub-divisions of 'dependent'. Since relationships are classified in just the same way as nodes, a typical dependency inherits by default from a number of higher-level dependencies; for example, in the sentence *He likes her* the relation between *likes* and *her* inherits from 'object', 'complement', 'valent', 'post-dependent' and 'dependent', each of which brings with it inheritable characteristics. (This relations are defined by the hierarchy shown in figure 19.4.)

Syntactic structure is primarily linear, so it is important to be able to indicate word order. A network has no inherent directionality, so linear order is shown as a separate relationship between earlier and later, by means of an arrow which points towards the earlier member of the pair; this arrow is labelled '<<'. (The linear order relationship has many other applications beside word order - they order points of time, so within language they are also used in the semantics of tense to relate the time of the verb's referent to the deictic time of its utterance.) Like any other relationships they can be overridden in the inheritance hierarchy, so it is easy to model the idea of a 'basic' and 'special' word order. For example, in English (a head-initial language) the basic word order puts words before their dependents, but enough dependents precede their heads to justify a general sub-type 'pre-dependent', of which 'subject' is a sub-type; so, exceptionally, a verb follows its subject. However there are also exceptions to this exception: an 'inverting' auxiliary verb reverts to the position before its subject.

This hierarchy is shown in figure 19.12. In words:

- A typical dependent follows its parent (the word on which it depends): *...likes her*.
- But a pre-dependent precedes its parent.
- Therefore a subject (one kind of pre-dependent) precedes its parent: *He likes ...*
- But the subject of an inverting auxiliary follows its parent: *Does he ...?*

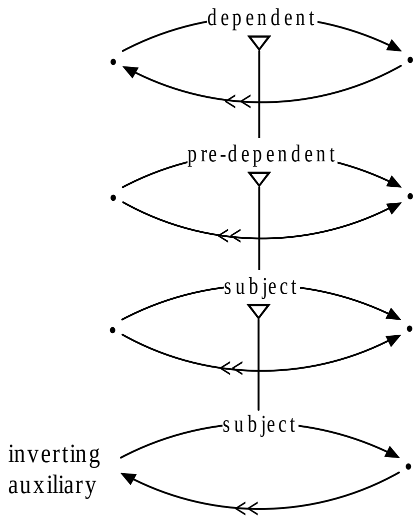


Figure 19.12: Default and exceptional word orders in English

A further source of flexibility in explaining word order is the fact that syntactic structure is embedded in a network theory, which (in principle) allows unrestricted links among nodes. This flexibility is in fact limited, but some multiple links are permitted. Not only may one word have more than one dependent, but one word may also have more than one parent. This is the source of most of the well-known complications in syntax such as raising, extraction and extraposition. Sentence (1) below contains examples of all three and shows how they can be explained in terms of a tangle-free 'surface' structure which is displayed above the words, supplemented by extra dependencies below the words. The Word Grammar analysis is summarised in Figure13 (which ignores all the isa links to the grammar network).

(1) It surprises me what she can do.

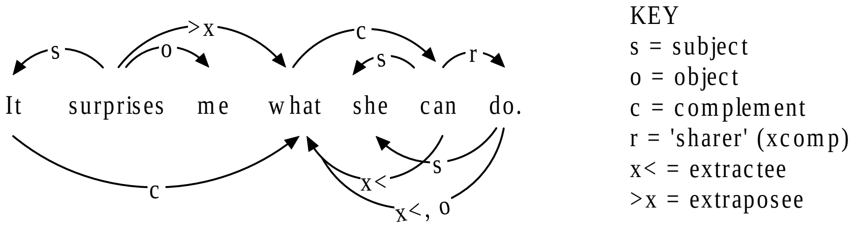


Figure 19.13: An example illustrating raising, extraction and extraposition

One of the attractions of this kind of grammar is that the structures combine concreteness (there are no word orders other than the surface one) with abstractness (the dependencies can show abstract relationships between non-adjacent words, and are generally in step with semantic relationships). This makes it relatively easy to teach at an introductory level, where it is possible to teach a grammatical system which can be applied to almost every word in any text (Hudson 1998). However at the same time it allows relatively sophisticated analyses of most of the familiar challenges for syntactic theory such as variable word order, co-ordination and Pied-piping.

11. Lexical semantics

According to Word Grammar, language is an area of our general conceptual network which includes words and everything that we know about them. This area has no natural boundaries, so there is no reason to distinguish between a Word's 'truly linguistic' meaning and the associated encyclopedic knowledge. For example, the sense of the word CAT is the concept Cat, the same concept that we use in dealing with cats in everyday life. It would be hard to justify an alternative analysis in which either there were two 'cat' concepts, one for language and one for the encyclopedia, or in which the characteristics of the Cat concept were divided into those which do belong to language and those which do not. This general philosophy has been applied in detail to the word CYCLE (Holmes, Hudson 2001).

In short, as in most other 'cognitive' theories of semantics, a word's meaning is defined by a 'frame of knowledge' (Fillmore 1985). In the case of Cat, the relevant frame includes the links between this concept and other concepts such as Pet, Mammal, Dog, Mouse, Fur, Milk and Miaowing. This frame of background knowledge is highly relevant to the understanding of language; for example, the link to Pet provides an easy interpretation for expressions like *our cat* in contrast with, say, *our mouse*.

Moreover any theory of language must make some attempt to formulate linking rules which map semantic relations to syntactic relations. For instance, we must at least be able to stipulate that with the verb HEAR, the hearer is identified by the subject, in contrast with SOUND which links it to the prepositional object as in *That sounds good to me*; and it would be even better if we could make these linkages follow from more general facts. Word Grammar has the advantage of syntactic and semantic structures that have very similar formal properties, so they should be easy to map onto one another. A syntactic structure consists of labelled dependencies between words, and a semantic structure consists of labelled dependencies between the concepts expressed by these words. Moreover, the labels reflect a hierarchical classification so each relationship can be identified at a number of different levels of generality. For example, in syntax 'object' isa 'complement', and similarly in semantics 'hearer' isa 'perceiver' which isa 'experiencer', and so on. In principle, then, Word

Grammar provides at least a good framework for exploring whatever linkage generalisations there are about the mapping from semantic relations to syntactic ones (Gisborne 2001).

Figure 19.14 shows sub-networks for the verbs HEAR and SOUND which illustrate this claim (Gisborne 1996). In words, the subject and object of HEAR define the hearer ('er') and hear-ee ('ee'), whereas SOUND has a more complex semantic structure. 'Sounding' is a kind of Judging, in which the evidence is an example of Hearing the individual to whom the judgement applies; for example, if John sounds nice to me, I base the judgement that John is nice on hearing John. (Another possibility is that my evidence is hearing something other than John; this meaning requires a slightly different analysis which may be combined with the one presented here.)

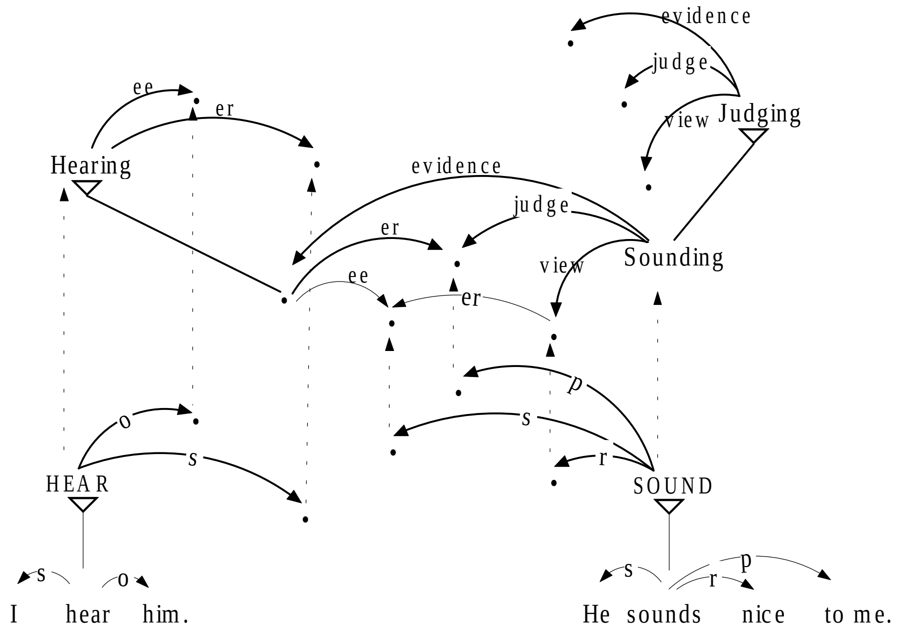


Figure 19.14: The semantics and syntax of the verbs SOUND and HEAR

The network approach allows analyses of very rich and complex areas of real-world knowledge such as the 'scenes', 'scripts' or 'frames' analysed in a variety of other frameworks (Barsalou 1992: 157; Sowa 1984; Luger, Stubblefield 1993: 368-386). It has the great advantage of avoiding all boundary problems which arise in theories which assume rigid 'frames' in which each item of information must be assigned to a single frame. For example, the concept Money belongs in part to the Commercial Transaction scene, but it also belongs to many other scenes - Work, Banking, Economics, Richness and many more. In a network analysis the one concept may be linked to all these other concepts at the same time.

12. Compositional semantics

Combining the meanings of individual words to make a composite 'sentence meaning' is quite easy given that:

- the words are related syntactically by word-word dependencies;
- some individual words have semantic structures which are linked to particular syntactic dependencies (as illustrated in figure 19.14).

However the network architecture has interesting consequences here as well, because a word's meaning changes when it is modified by a dependent. For example, the meaning of *cat* in the phrase *a small cat* is 'small cat' rather than simply Cat; and the particular cat is another concept again. Since these are distinct concepts they must be represented by distinct nodes, with the consequence that the semantic structure contains a separate node for every phrase that would have been recognised in a conventional phrase structure analysis: one for *cat*, another for *small cat* and a third for *a small cat*. This pattern is called 'semantic phrasing' (Hudson 1990: 146-51). In most cases the relation between the nodes is 'isa': the particular cat isa Small cat, which in turn isa Cat. Well-known exceptions to the isa relationship include the effect of adding FAKE (e.g. Fake diamonds) and NOT; metaphors also break the isa link. In the normal case, the word's referent (e.g. our mental representation of the particular cat) isa all its other meanings, which we may call collectively its senses.

Figure 19.15 illustrates the kinds of semantic structures which are stored permanently in the lexico-grammar, and which are the building blocks out of which sentence structures are constructed. In words:

- The word A shares the same referent as its complement noun. (A complete analysis also shows that this referent is 'indefinite' and countable.)
- SMALL, when used as a pre-adjunct ('a<') of a noun, modifies the latter's sense by specifying a value for its 'size' which is less than the default value for the relevant prototype.
- CAT means Cat.
- MIAOW means Miaowing, which has a 'miaower' ('er') which isa Cat; this concept is shared by the words CAT and MIAOW.

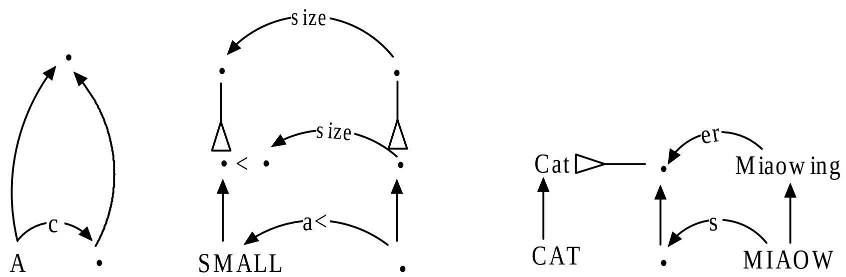


Figure 19.15: The syntax and semantics of the words A, SMALL, CAT and MIAOW.

Given an accurate syntactic analysis, these meanings combine into the compositional semantic structure shown in figure 19.16, in which there are separate labelled nodes for the concepts 'Cat', 'Small cat', 'A small cat', 'A small cat miaowing' and 'A small cat miaowed' (the particular instance of miaowing referred to here), each of which is a distinct semantic element corresponding to a 'phrase' in the syntax. And yet this semantic structure is built in a very simple way from a completely flat syntactic structure.

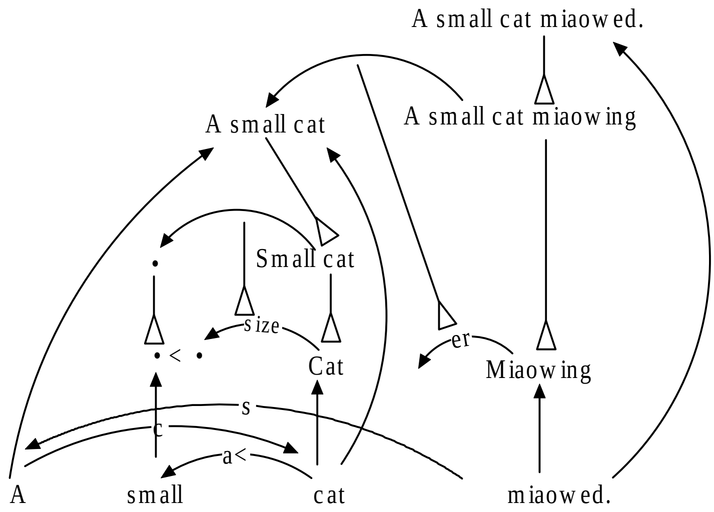


Figure 19.16: The syntactic and semantic structure of *A small cat miaowed*.

Semantic phrasing also helps with some of the standard challenges of logic. For example, take a simple sentence such as *John kissed a girl*. The semantic structure contains one node which shows the modifying effect of the object, and another for the subject:

- Kissing a girl,
- John kissing a girl.

The crucial question is exactly how these are related to each other. Clearly the second isa the first, but what about their respective 'kiss-ee' arguments? One possibility is that they too are in an isa relationship, rather than simple identity. This is the analysis shown in figure 19.17. The part to pay attention to here is the 'isa' link between the variables y and x, showing that the girl of 'John kissing a girl' isa the one in 'Kissing a girl'.

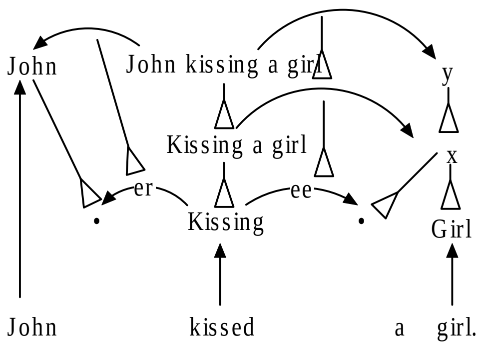


Figure 19.17: Semantic structure for *John kissed a girl*.

In this sentence the effect on the logic is exactly the same as if the same girl had been involved in both cases because y has no characteristics other than those which it automatically inherits from x. But suppose there were several different examples of 'Kissing a girl'. In that case the analysis in figure 19.17 would allow the girl to be either the same or different. For example, take sentence (1).

(1) John kissed a girl, and Bill did too.

In this sentence, *did too* refers anaphorically to *kissed a girl*, but allows both interpretations: either Bill kissed the same girl as Bill, or he kissed a different one. Both interpretations are compatible with the analysis in which 'Bill kissing a girl' isa

'Kissing a girl' and the girl in the former isa the one in the latter. In short the possibility of 'sloppy identity' follows from the logic of default inheritance combined with semantic phrasing.

The same assumption about identity explains the ambiguity of coordinate structures such as (2).

(2) John and Bill each kissed a girl.

Because of *each*, this has to refer to two distinct instances of 'Kissing a girl', one performed by John and the other by Bill; but if each of the girls isa the one implicated in 'Kissing a girl' the girls may be either the same or different. In other words, contrary to standard logical analyses in terms of scope, the sentence is not in fact ambiguous, but simply vague. This seems right because if we add a third male, Mike, the sentence would be compatible with a scene in which John and Bill kissed the same girl and Mike kissed a different one - an interpretation which would be hard to represent in terms of the predicate calculus. A similar analysis applies to an example such as (3).

(3) Every boy kissed a girl.

As in the previous examples, this is vague rather than ambiguous. There is a distinct instance of 'Kissing a girl' for each boy, but the girl in each of these cases simply isa Girl and might be either the same or different.

Word Grammar needs a great deal more work in this area but some foundations are already available (Hudson 1984, 131-210; Hudson 1990, 123-165).

13. Sociolinguistics

Word Grammar is one of the few theories of language structure in which there is any provision for 'sociolinguistic information' - the kind of knowledge that allows us to interpret utterances in terms of social categories such as speaker types and interaction types. Thanks to recent work in sociolinguistics we know a great deal about the ways in which people classify speakers, in terms of geography, social class, age, sex and so on, and in terms of speaking situations, as chatting, teaching, greeting, joking, etc. (Hudson 1996). Classification clearly presupposes knowledge ('competence'), just like the rest of language, so any cognitive theory of language must accommodate it in some way.

The Word Grammar solution is to recognise that words are actions. This is much easier to accept if we think of words as spoken rather than written, and if we compare them with recurrent actions such as cleaning one's teeth, for which we have permanent stored representations. We know how and when to clean our teeth in much the same way that we know how and when to use a Word, and we can distinguish the stored 'type' from all of the particular 'tokens' of it that we perform. The Word Grammar claim, then, is that a word type is like the action type 'Cleaning one's teeth' - a stored concept for a particular kind of action (the action of saying the relevant sounds) for a particular purpose and in a particular kind of social context.

Now if Word isa Action, it must inherit the latter's characteristics, one of which is that it has an actor; in the case of a word, the actor of course is the speaker, so this analysis immediately allows us to represent the speaker in the analysis of a Word. (This is also helpful in handling deictic semantics such as the reference of the word ME - the referent of ME is its actor/speaker.) Furthermore, we can classify actions according to the kinds of situation in which they are socially appropriate - for example, spitting is fine when cleaning our teeth but not when eating. In the same way we can classify words, or word combinations, according to the social situations in which they are appropriate. Figure 19.18 shows how the role 'speaker' can be exploited in the language network.

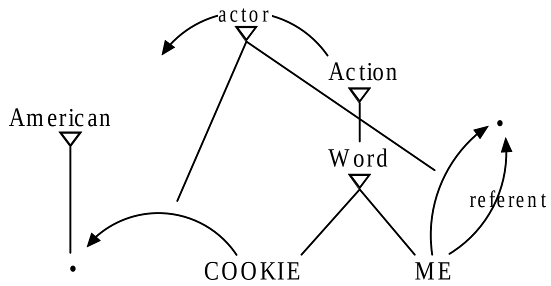


Figure 19.18: Constraints on the speaker of COOKIE and ME

The Word Grammar work on sociolinguistics is mostly programmatic rather than thorough, but there is a small amount of Word Grammar work on quantitative sociolinguistics (Hudson 1996: 243-257, Hudson 1997a; Hudson 1997b).

14. Processing

Any theory of how we store and organise language as knowledge must also be compatible with some theory of how we use this knowledge as speakers, listeners, writers and readers. I have already suggested various ways in which the network theory of knowledge fits what we know about these various kinds of processing. The various claims made earlier are summarised here:

- A network can be activated, and when one part is activated the activation naturally spreads to neighbouring parts. We know that spreading activation is a reality, as evidenced by priming in perception and by speech errors in production. It is easy to model spreading activation if the language itself is modelled as a network.
- A Word Grammar network is built round a number of 'isa' hierarchies which allow default inheritance. This explains a number of processing effects - how we use observable word forms to derive information about unobservable meaning, how we generalise to unfamiliar cases, how we cope with exceptions and even with deviant input.
- The possibility of exceptions and deviant input which follows from default inheritance explains why processing requires a global Best Fit Principle rather than more rigorous local tests for well-formedness; for example, when pushed strongly by context we may overlook a gross misspelling or mispronunciation.
- Returning to spreading activation, this helps to explain how we cope with the potentially fatal problems of both default inheritance and the Best Fit Principle, both of which in principle require us to search the whole of our knowledge base for more specific overriding facts or better global fits. If we assume that all relevant nodes are already active, then the search for alternatives can focus on these and ignore the rest of the database.

15. Learning

Lastly, the Word Grammar theory of language structure is what would be expected according to the Cognitive Linguistics view that language is learned on the basis of usage (Barlow, Kemmer 2000). If this view is correct, then we should expect:

- degrees of entrenchment, as recognised in section 3, whereby the more often we activate a particular link, the more deeply entrenched it is;

- lexical detail as well as (and prior to) more schematic generalisations across lexemes;
- linguistic categories of all kinds (words, syntactic patterns, phonemes) which are sensitive to features of the non-verbal context.

All these characteristics are supported by a great deal of empirical evidence from studies of child language (Ellis 2002; Lieven, Pine, Baldwin 1997; Pine, Lieven, Rowland 1998; Tomasello 2000, 2003).

We have already shown how degrees of entrenchment can be attached to either nodes or links in a network, but the network postulate also helps to explain the other characteristics. If all knowledge is indeed a single integrated network, then this network must include knowledge of the tokens that we analyse as well as the stored knowledge that we apply to them. We have assumed that the relationship between the two is the 'isa' relationship, so each word token isa some word type which in turn isa various more general types. If this is correct, then learning is rather simple: it involves no more than the conversion of token nodes into type nodes. That is, instead of allowing a token node to die away for lack of activation, we activate it sufficiently to keep it alive for future use as a type for classifying other tokens. Tokens are the ultimate in lexical specificity, so this process explains why children start with lexically specific patterns before inducing generalisations; and the fact that tokens are always contextualised explains why we can learn contextual ('sociolinguistic') information about linguistic items.

Finally, a rather different feature of Word Grammar turns out to be highly relevant to this account of language learning. This is the use of dependencies in syntax. Unlike phrase structure analyses, dependency analysis allows us to measure the distance between a word and the word on which it depends - the 'dependency distance' (Hiranuma 1999; Hiranuma 2001). It turns out that in casual speech dependency distance is zero for most words - typically 70% or more of words are next to the word on which they depend (Collins 1996; Pake 1998). Moreover, every English dependency pattern may be found between adjacent words. These two facts mean that a child can learn syntax very easily by paying attention to nothing but adjacent word pairs, and ignoring the 30% of patterns which do not recur.

This article has summarised the main features of Word Grammar as of mid 2003, but the theory is continuously evolving. More up-to-date information may be found on the Word Grammar web site, http://www.phon.ucl.ac.uk/home/dick/wg.htm.

Reference List

Aitchison, Jean. 1997. *The language web*. Cambridge: Cambridge University Press.

Barabási, Albert-László. 2003. Linked: How everything is connected to everything else and what it means for business, science and everyday life. London: Penguin.

Barlow, Michael and Suzanne Kemmer. 2000. *Usage Based Models of Language*. Stanford: CSLI.

Barsalou, Lawrence. 1992. *Cognitive Psychology. An overview for cognitive scientists.* Hillsdale, NJ: Erlbaum.

Bock, Kathryn and Zenzi Griffin. 2000. The persistence of structural priming: Transient activation or implicit learning? *Journal of Experimental Psychology -General* 129: 177-192.

Bybee, Joan. 1995. Regular Morphology and the Lexicon. *Language and Cognitive Processes* 10: 425-455.

Bybee, Joan. 2001. *Phonology and language use*. Cambridge: Cambridge University Press.

Chomsky, Noam. 1986. *Knowledge of Language. Its nature, origin and use*. New York: Praeger.

Collins, Michael. 1996. A new statistical parser based on bigram lexical dependencies. *Proceedings of the Association for Computational Linguistics*.

Creider, Chet and Richard Hudson. 1999. Inflectional Morphology in Word Grammar. *Lingua* 107: 163-187.

Ellis, Nick. 2002. Frequency effects in language processing: a review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition* 24.

Fillmore, Charles. 1985. Frames and the semantics of understanding. *Quaderni di Semantica* 6: 223-253.

Fodor, Jerry. 1983. *The Modularity of the Mind.* Cambridge, MA: MIT Press.

Garfield, Jay, ed. 1987. *Modularity in Knowledge Representation and Natural-Language Understanding*. Cambridge, MA: MIT Press.

Gisborne, Nikolas. 1996 *English Perception Verbs.* PhD dissertation, University of London.

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: University of Chicago Press.

Halle, Morris and Sylvain Bromberger. 1989. Why phonology is different. *Linguistic Inquiry* 20: 51-70.

Halliday, Michael. 1961. Categories of the theory of grammar. *Word* 17: 241-292.

Harley, Trevor. 1995. *The Psychology of Language*. Hove: Psychology Press.

Hiranuma, So. 1999. Syntactic Difficulty in English and Japanese: a Textual Study. *UCL Working Papers in Linguistics* 11: 309-322.

Hiranuma, So. 2001. *Measuring the syntactic difficulty of Japanese sentences.* PhD dissertation, University of London.

Holmes, Jasper and Richard Hudson. forthcoming. Constructions in Word Grammar. In Jan-Ola Östman & Mirjam Fried, eds., *Construction Grammar(s): Cognitive dimensions*. Amsterdam: Benjamins.

Hudson, Richard. 1971. *English Complex Sentences. An introduction to systemic grammar.* Amsterdam: North Holland.

Hudson, Richard. 1984. *Word Grammar.* Oxford: Blackwell.

Hudson, Richard. 1990. *English Word Grammar.* Oxford: Blackwell.

Hudson, Richard. 1996. *Sociolinguistics, 2nd edition*. Cambridge: Cambridge University Press.

Hudson, Richard. 1997a. Inherent variability and linguistic theory. *Cognitive Linguistics* 8: 73-108.

Hudson, Richard. 1997b. The rise of auxiliary *DO*: Verb-non-raising or category-strengthening? *Transactions of the Philological Society* 95: 41-72.

Hudson, Richard. 1998. *English Grammar.* London: Routledge.

Hudson, Richard. 1999. Subject-verb agreement in English. *English Language and Linguistics* 3: 173-207.

Hudson, Richard. 2000a. *I amn't. *Language* 76: 297-323.

Hudson, Richard. 2000b. Discontinuity. *Traitement Automatique des Langues.* 41: 15-56.

Hudson, Richard. 2000c. Grammar Without Functional Categories. In Robert Borsley, ed., *The Nature and Function of Syntactic Categories* 7-35. New York: Academic Press.

Hudson, Richard. 2003a. Trouble on the left periphery. *Lingua* 113: 607-642.

Hudson, Richard. 2003b. Gerunds without phrase structure. *Natural Language and Linguistic Theory* 21: 579-615.

Jackendoff, Ray. 1997. *The Architecture of the Language Faculty*. Cambridge, MA: MIT Press.

Lamb, Sydney. 1966. *Outline of Stratificational Grammar*. Washington, DC: Georgetown University Press.

Lamb, Sydney. 1998. *Pathways of the Brain. The neurocognitive basis of language.* Amsterdam: Benjamins.

Langacker, Ronald. 1998. Conceptualization, symbolization and grammar. In Michael Tomasello, ed., *The New Psychology of Language: Cognitive and functional approaches to language structure* 1-39. Mahwah, NJ: Erlbaum.

Langacker, Ronald. 2000. A dynamic usage-based model. In Michael Barlow and Suzanne Kemmer, eds.,*Usage-based Models of Language* 1-63. Stanford: CSLI.

Lieven, Elena, Julian Pine, and Gillian Baldwin. 1997. Lexically-based learning and early grammatical development. *Journal of Child Language* 24: 187-219.

Luger, George and William Stubblefield. 1993. *Artificial Intelligence. Structures and strategies for complex problem solving* . New York: Benjamin Cummings.

Pake, James. 1998. *The Marker Hypothesis. A constructivist theory of language acquisition.* PhD dissertation, University of Edinburgh.

Percival, Keith. (1976). On the historical source of immediate constituent analysis. In McCawley, J. D.(ed.), *Notes From the Linguistic Underground.* London: Academic Press.

Percival, Keith. (1990). Reflections on the History of Dependency Notions in Linguistics. *Historiographia Linguistica* 17: 29-47.

Pine, Julian, Elena Lieven, and Caroline Rowland. 1998. Comparing different models of the development of the English verb category. *Linguistics* 36: 807-830.

Radford, Andrew. 1997. *Syntactic Theory and the Structure of English.A minimalist approach*. Cambridge: Cambridge University Press.

Reisberg, Daniel. 1997. *Cognition. Exploring the science of mind.* New York: Norton.

Robins, Robert. 1959/2001. In defence of WP. *Transactions of the Philological Society* 99: 114-144.

Roelofs, Ardi. 1996. Serial order in planning the production of successive morphemes of a Word. *Journal of Memory and Language* 35: 854-876.

Rosta, Andrew. 1997. *English Syntax and Word Grammar Theory* PhD dissertation, University of London.

Sadock, Jerrold. 1991. *Autolexical Syntax: A theory of parallel grammatical representations.* Chicago: University of Chicago Press.

Sowa, John. 1984. *Conceptual Structures: Information Processing in Mind and Machine.* Reading, MA: Addison-Wesley.

Sperber, Dan and Deirdre Wilson. 1995. *Relevance. Communication and cognition.* Oxford: Blackwell.

Sweetser, Eve. 1987. The definition of lie. In David Holland and Nigel Quinn, eds., *Cultural Models in Language and Thought* 43-66. Cambridge: Cambridge University Press.

Tomasello, Michael. 2000. Do young children have adult syntactic competence? *Cognition* 74: 209-253.

Tomasello, Michael. 2003. *Constructing a language: a usage-based theory of language acquisition.* Cambridge, MA: Harvard University Press.

Winograd, Terry. 1976. Towards a procedural understanding of semantics. *Revue Internationale de Philosophie* 30: 260-303.

---

[i] I should like to thank Joe Hilferty and the editors for helpful comments on an earlier version.