Hudson, Richard. 2000. Language as a cognitive network. In Hanne Gram Simonsen & Rolf Theil Endresen (eds.), A Cognitive Approach to the Verb. Morphological and Constructional Perspectives, 49–72. Berlin: Mouton de Gruyter. http://www.phon.ucl.ac.uk/home/dick/net.htm.

Language as a cognitive network

Richard Hudson[1]

Abstract
The paper develops the leading idea of Word Grammar and other "cognitive" theories of language, which is that language is a network. It reviews some of the consequences of this view: spreading activation, effects of conceptual distance, default inheritance, the unity of grammar and lexicon and, more generally, non-modularity; the unity of permanent and temporary representations, degrees of accessibility and binary relations. It then shows briefly how these ideas apply to two specific areas of language analysis: the contrast between polysemy and homonymy, and the treatment of regular and irregular morphology. The last section discusses Pinker's contrast between "mentalese" and "connectoplasm", and argues that the networks defined in this paper have all the symbolic qualities of mentalese, so maybe the mind uses "networks all the way down".

Keywords: network, language, polysemy, morphology, activation, modularity

1 Background: Networks in Cognitive Linguistics and Word Grammar

The main focus of this paper is one aspect of the approach to language which has come to be called "Cognitive Linguistics" (Lakoff 1987, Langacker 1987, 1990, Taylor 1989), namely the widespread assumption that the whole of language is a network, a collection of units linked pairwise by relations. These units are concepts - concepts such as "word", "the word *word*" or "the phoneme /p/" - and the leading claim of Cognitive Linguistics is that the concepts of language are organised in the same way as all other concepts. In short, "knowledge of language is knowledge" (Goldberg 1995:5). General knowledge is widely believed to be structured as a network, so this view allows language to be integrated with other kinds of knowledge.

The network view contrasts most clearly with the view that knowledge is divided into distinct modules, each with its own formal properties and internal structure and with very limited links to other modules (Fodor 1983, Chomsky 1986, Pinker 1994, 1997). As far as language is concerned, the lexicon and the grammar form distinct modules, as do phonology, semantics and so on; and in an extreme version, each lexical item is held in a distinct "lexical entry" which is a kind of mini-module.

Cognitive Linguistics exists as a rather general ideology about language, but it also includes three well-developed theories of language structure: Cognitive Grammar (Langacker 1987, 1991, 1994), Construction Grammar (Goldberg 1995) and my own theory, "**Word Grammar**" (**WG**[1]). Networks have played a central role in this theory since its earliest days, as shown by the following quotation:

> "A language is a network of entities related by propositions." (Hudson 1984:1)

However, its historical roots are rather different from those of the other theories. It

[1] University College London; dick@ling.ucl.ac.uk

was influenced heavily by Chomsky's work, and especially by his stress on generativity, but it does not, in any way, grow out of Transformational Grammar. (In personal terms, I have never been a Chomskyan - witness Hudson (1976).) The main theories of grammar that contributed to it are also outside the Chomskyan main-stream: Systemic Functional Grammar (Halliday 1985, Martin 1992, Hudson 1971), Stratificational Grammar (Lamb 1966, Bennett 1992) and Dependency Grammar (Tesnière 1959/66). As will become apparent below, some of the leading ideas come directly from Cognitive Science.

The undoubted similarities between WG and the other "cognitive" theories (Cognitive Grammar and Construction Grammar) are due to independent evolution. This in itself is encouraging for those of us who would like to believe that our theories are driven by the facts rather than by our personal histories. We shall see some differences between WG and the other theories, but the most obvious differences are matters of notation and the substantive differences are relatively unimportant compared with the shared stress on networks. We shall now consider some of the consequences of building a theory of language in terms of networks.

2 Networks

2.1 Networks and activation

For a cognitive linguist, one of the attractions of applying networks to language structure is that cognitive psychologists often model other parts of cognition as a network. One of the most widely accepted ideas in this field is "that memory connections provide the basic building blocks through which our knowledge is represented in memory ... all of knowledge is represented via a sprawling network of these connections, a vast set of associations." (Reisberg 1997:257-8). If language is like other parts of cognition, then we might expect at least some parts of language to constitute a network; and if some parts are a network, the most parsimonious theory is that they all are. This is the basis for the WG claim quoted above: "a language is a network of entities related by propositions." Cognitive Grammar and Construction Grammar also present language as a network of units (Langacker 1990:2, 266, Goldberg 1995:5).

What does this claim mean, and how is it distinctive? The network view has a great many interesting and important consequences, but I shall highlight just a selection. First, knowledge is declarative, rather than procedures which have an "input" and an "output". We know relationships, not "rules" for "doing" things. For example, we know that a past-tense verb has a suffix after its stem, and that the stem and suffix together comprise the whole of the verb; these are relationships among the concepts "past-tense verb", "its suffix", "its stem" and "its whole" which are diagrammed (using WG notation) in Figure 1. We do not know "how to form a past-tense verb" - such knowledge cannot be accommodated, as such, in a network. A network is like a map, which is simply static; it is not like a set of instructions for getting from A to B.

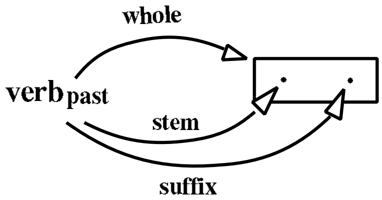

Figure 1

On the other hand, because of its declarative nature, the network may be used like a map as a guide to activity. When our goal is to find a pronunciation we can move from "past-tense verb" to the pattern "stem + suffix" by following these links in the network, and when our goal is to find a meaning we can move in the opposite direction. This is possible in a network because links may be activated like the links on a circuit board. A widely accepted model in cognitive psychology is **spreading activation**, whereby activation of one node spreads (as weaker activation) to the nodes to which it is directly attached. Spreading activation can only be assumed in a network; it makes no sense if knowledge is envisaged as a collection of rules or separate lexical items.

It is perhaps in this sense that Langacker states that "linguistic knowledge is procedural rather than declarative" (ibid:15). Activation is central to Cognitive Grammar, and the "structured inventory of conventional linguistic units" (ibid) which are activated is closely analogous to the network of concepts in WG.

2.2 Conceptual distance and the networked lexicon

The conceptual distance between two concepts can be measured as the number of intervening links - a notion which would be meaningless if the same information was presented as a list of rules or propositions. We shall explore this consequence below, but meanwhile it should be noticed that distance is relevant because of spreading activation - the further apart two concepts are, the less likely they are to activate one another. The classic experiments on **semantic priming** (Reisberg 1997:265-7, Swinney 1979) show that activation can spread across several links, so it does not require strict adjacency. For example, when presented with the word *ink*, subjects take less time to decide that it is an existing English word when it has been primed by an immediately preceding word than when it has not; and it is primed by *pen*, but not by *king*. Now the point to notice is that the link between the words *ink* and *pen* lies in their meaning: (the word) *ink* - (the thing) "ink" - "pen" - *pen*. This involves two intervening concepts, and perhaps more (depending on how direct the link between "ink" and "pen" is). Of course the conclusion is not that *ink* is related to *pen* but not at all related to *king*; after all, in a network everything is ultimately related to everything else. The difference lies in the relative distance between the concepts, which is shown in Figure 2.

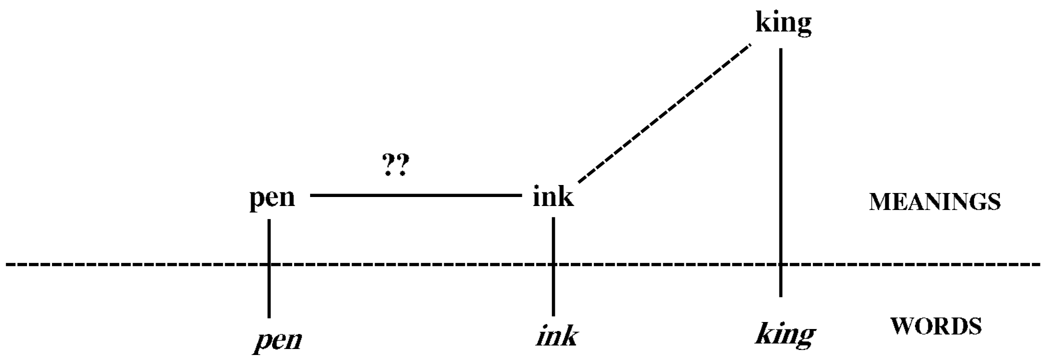

Figure 2

The same kind of spreading activation which leads to semantic priming also produces well-known effects in the recognition of words. These effects are the basis of the "**cohort**" model of speech recognition (Marslen-Wilson 1980, Marslen-Wilson and Tyler 1980, 1981) according to which the incoming sound-segments activate a decreasing cohort of candidate words: /e/ activates all the words that begin with /e/, then /l/ reduces the cohort to those that have initial /el/, and so on, until only one candidate remains - which is why a word like *elephant* can be recognised long before the end.

This evidence suggests, then, that when we hear a word it is embedded in a network of phonological and semantic links to other words. If this is so in speech perception, we should expect the same to be true of production - as it seems indeed to be, to judge by the evidence of speech errors. When a target word is replaced by another, the two are usually similar either in phonology or in meaning (Fromkin 1988), suggesting that these characteristics are "networked" (Emmory and Fromkin 1988:145).

It seems almost beyond dispute that the links from words to meanings and sounds involve a network of links between individual words (or lexemes) and individual semantic and phonological patterns - a far cry from the earlier view of the lexicon as a list of discrete lexical entries. It is less important to decide whether these links are just relationships or "rules" (e.g. the "correspondence rules" of Jackendoff 1996); indeed it is hard to know what the basis of such a distinction might be.

2.3 Default inheritance

Large parts of the network are organised hierarchically on the basis of the most important relationship of all, the "**isa**" relationship (which is widely recognised in Artificial Intelligence - see Reisberg 1997:280) - Langacker's "categorization" (1990:266), Goldberg's "instance link" (1995:79). The importance of this relationship lies in the fact that it allows "generalisation by inheritance". If A isa person, and a typical person has a heart, then A must have a heart as well; so even if it is true, this fact can be stored just once, in connection with the general category "person", and inherited each time it is needed. A classic experiment by Collins and Quillian (1969) showed that it takes longer to verify a fact which can be inherited than one which cannot (Reisberg 1997:267-70). For example, subjects take longer to verify that a canary can fly than that it can sing, because the ability to fly is inherited from "bird" but the ability to sing is a fact about canaries; and it takes even longer to verify that a canary has skin, because this is inherited via "bird" from "animal". In short, the time taken for verification depends on the conceptual distance between the inheriting node and the source of the information.

Inheritance is clearly one of the effects of activation which is controlled by conceptual distance. One particular consequence is that properties are only inherited by default, i.e. in the absence of a more specific specification to the contrary. In short, the process is more accurately described as **default inheritance** - "normal mode inheritance" (Goldberg 1995:73), "schematicity" with either "elaboration" or "extension" (Langacker 1990:266-7).

The overriding of defaults follows from the network model because the inherited property is always the one which involves the shorter route - i.e. the least conceptual distance. Take the canary example just quoted. The ability to sing is directly related (in the network) to "canary", giving a distance of just one link, whereas the ability to fly is separated from "canary" by two links: from "canary" to "bird", and from "bird" to "can fly". As we saw, retrieval is faster over shorter distances. Suppose now that the bird species in question was "ostrich". Unlike the typical bird, the typical ostrich cannot fly, so this fact must be stored directly in relation to "ostrich". If we now assume for simplicity that the link is labelled "flies?", with "yes" and "no" as possible values, it is clear that the value "no" will be retrieved before the value "yes". The relevant network is shown in Figure 3, alongside a similar network for verb morphology which shows why irregular morphology takes priority over regular. In this case, the relationship whose default value is overridden is "whole", which gives the word's fully-inflected form.

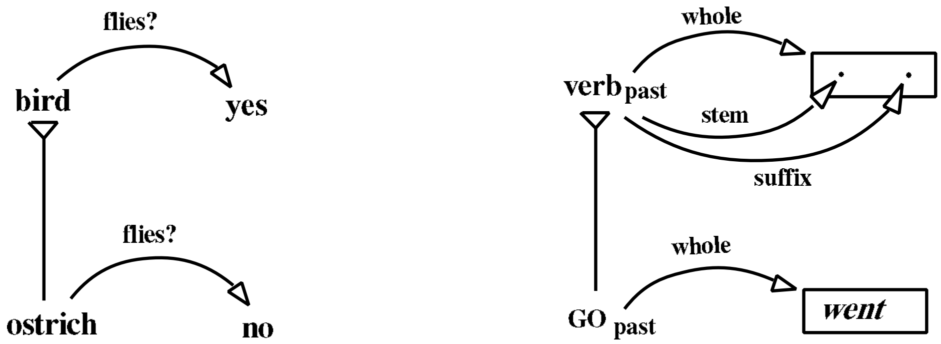

Figure 3

2.4 The lexicogrammar continuum and "temporary competence" in performance

If generalisations are made by default inheritance down an isa hierarchy, the facts which other theories express by means of general "rules of grammar" will be expressed as facts about relatively general categories near the top of the hierarchy. As in other cognitive theories, this leaves no place for a boundary between "the lexicon" and the rest of the grammar (Langacker 1990:102, Goldberg 1995:7); so if "the lexicon" is a network, as argued in 2.2, the same must be true of "the rules of grammar". Moreover, WG syntax uses word-word dependencies rather than phrase structure, so the grammar generates just words and their dependencies. This means that all higher-level generalisations apply to single-word categories - to the category "word" itself, or word-classes such as "noun" or "auxiliary verb".

The result is that all our linguistic competence is organised around an **isa hierarchy of word-types**, with "word" at the top and individual lexemes and word-forms at the bottom - the "lexicogrammar" (Halliday 1985). Indeed the hierarchy is continued even further down during **performance** by the addition of two extra kinds of word. First, there are word-types which are inherited rather than stored. For example, it is generally agreed that at least some regularly inflected words are not stored; so a newly invented word obviously cannot have been stored by the hearer, and yet we can all understand inflected forms like *wugs*. Presumably the same is true at least for inflected forms of regular and rare words like *disambiguate*. In order to understand the form *disambiguated* we must assign it a new temporary concept which isa the lexeme *DISAMBIGUATE* and "past-tense verb" but which is also distinct from the word token which we are understanding (whose properties need not be exactly the same - it could be mis-spelt, for example). In WG notation, this word-type is called $DISAMBIGUATE_{past}$.

The second kind of temporary addition to the network is the actual **word token** as recognised in processing. Suppose I am reading the following sentence:

(2) Pat shuffled uphill and Jo shufled [sic] downhill.

This contains seven word tokens, but two are tokens of the same type: $SHUFFLE_{past}$ (a concept which is unlikely to be stored). The WG name for the tokens is based solely on position, so they are called words 2 and 6, "w2" and "w6", giving three temporary concepts: $SHUFFLE_{past}$ and w2 and w6, each of which "isa" the former. The isa relationship allows mismatches, so this analysis is compatible with the mis-spelling shown. The hierarchy needed for these two tokens of *shuffled* is shown in Figure 4.

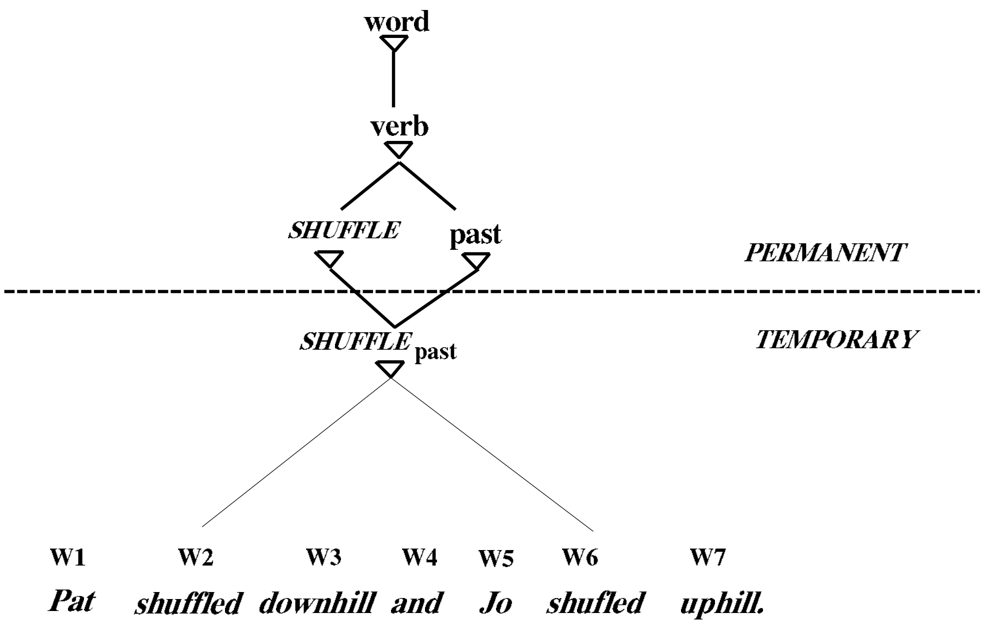

Figure 4

In this view of language, then, there is no boundary between grammar and lexicon, but nor is there any formal distinction between competence and the structures created in performance. Performance is the production of mental representations which "isa" those of competence, and we build our competence by permanently storing some representations that we produce during performance (depending on factors such as frequency, salience and so on). In short, WG is a "usage-based" theory, just like the other two theories (Langacker 1990:264, Goldberg 1995:133).

2.5 Non-modularity

More generally, there can be **no modularity**, in the sense of Chomsky (1986) or Fodor (1983). A network has no natural boundaries, which means that the network must have the same fundamental formal properties throughout. For instance, it could not allow spreading activation, or default inheritance, in some areas but not in others; nor could it accommodate a phonology which was procedural and a syntax which was declarative.

This is not to say that the network is completely uniform throughout. On the contrary, any cognitive network will reflect the boundaries of "nature", so that similar experiences will be more closely related to one another than to more different experiences, producing a pattern of "**subnetworks**" with more or less vague boundaries - one network for our family members, another for colleagues, a much larger one for people which includes both, another for telephone numbers, and so on. These subnetworks are by no means isolated from one another - for example, every telephone number is the number of some person, so numerous links go from individual numbers in the telephone-number subnetwork to individual people in the people subnetwork.

The same must surely be true of language, especially given the importance of isa hierarchies for bringing together units that are similar. For example, the relationships of syntax comprise a hierarchy of dependency-types (e.g. "subject", "complement", "adjunct") while those of morphology belong to a different hierarchy which includes "stem", "suffix" and "whole"; there is no direct link between these two hierarchies, and they probably each define a different subnetwork of the total grammar. Words belong to numerous overlapping subnetworks, but these need not correspond to the traditional divisions into linguistic modules. For example, brain damage can impair very specific areas of vocabulary such as all words which name things found inside the house (Pinker 1994:314); no linguistic theory recognises a module for such words or their meanings, but such cases are to be expected if knowledge is indeed clustered into subnetworks as suggested above. Given our ability to see connections in the world, knowledge must be compartmentalised to the extent that the world itself is.

2.6 Accessibility

Another well-established finding of psychology (and of ordinary experience) is that some concepts are more accessible than others. This has been particularly well documented for word retrieval; for example, accessibility depends on frequency and recency (Yelland 1994:5012).

These notions can surely be accommodated in some way in a network model, but they again make very little sense in terms of rules and lexical entries.

However, in a network-based theory we have to answer an important question: is it the nodes themselves, or their links, that vary in accessibility? In Cognitive Grammar nodes (e.g. structural schemas) are described as having various degrees of "entrenchment" (Langacker 1987:59), but my own view is that at least some variation applies to the links between nodes, with some links being more easily activated than others. For example, when we hear a familiar piece of music it is often hard to recall its composer, regardless of how "entrenched" that composer may be; what is inaccessible is not the composer, as a person, but the link from the tune to that person. This is easy to explain in terms of a network in which each connection between two nodes has a "weight" which decides how much activation it passes from one node to the other (Luger and Stubblefield 1993:517). If this is the explanation for all kinds of quantitative variation, then entrenchment will have been replaced by weighted links, but of course it may turn out that in some sense both nodes and links need to be weighted. This is clearly an important question for research.

There is an interesting connection between these ideas and the findings of quantitative dialectology. This work (largely led by William Labov) has uncovered a mass of quantitative links between speakers" social characteristics (sex, age, social class and so on) and their use of selected linguistic variables (e.g. the pronunciation of a particular phoneme or word). These findings have so far not been successfully integrated into any main-stream theory of language structure, but the network approach promises to be more helpful. Suppose a speaker S alternates between overt /r/ and no /r/ in words like *fourth* and *floor*, but regularly uses /r/ in about 35% of all tokens where /r/ is possible - a very typical finding of Labovian research (Hudson 1996, chapter 5). This behaviour must be based on S's knowledge of the local social structures, so let's assume that S believes /r/ is typically used by members of social group G. We now have a simple sub-network in S's mind which links the following concepts: speaker S, group G and pattern /r/. We can now explain S's behaviour as the result of the weightings of these links - for example, a weighting of 35% on the S-G link and 100% on the /r/-G link. It remains to be seen whether it is possible to build a single network model of this kind which integrates the figures for a larger number of linguistic variables (Hudson 1996:248-52).

2.7 Relationships

A concept's characteristics - its "definition" - all involve its relationships to other concepts, so there are **no "one-place predicates"**. This follows from the claim that a network expresses all of knowledge, and not just those bits of knowledge that involve traditional two-place predicates. This conclusion applies of course throughout the grammar, but its consequences are probably most controversial in semantics.

In traditional logic, a verb such as *yawn* is a one-place predicate, not a relationship; but in network analysis it must be a relationship between a person - the "yawn-er" - and another concept, the "act of yawning". Similarly, traditional logic allows a general category such as "linguist" to act as a predicate with a single argument, so *He is a linguist* expresses a proposition based on a one-place predicate. In a network, however, it involves two distinct concepts linked by the "isa" relationship.

Adjectives and adverbs raise rather special problems because there is no well-developed WG analysis of them, but in principle the same remarks apply to them. Thus an adjective like *bald* cannot be treated as a one-place predicate but defines a relationship. In this case the argument is supplied by the noun to which the subject is applied - e.g. by *linguist* in *bald linguist* - and the value is some degree of baldness, namely a degree which is above average which we can call simply "bald". Similarly for adverbs like *loudly*, whose argument is the verb to which they are applied, as adjuncts. Take the following example:

(3) The bald linguist yawned loudly.

By assumption the semantic structure of this sentence involves the pairs of concepts and relationships shown in Table 1.

argument	value	relationship
x (= the yawner)	linguist	isa
y (= the yawning)	yawning	isa
y	yawn-er	x
x	baldness	bald
y	loudness	loud

Table 1

These relationships are shown in WG notation in Figure 5, which is a partial structure for the sentence. It does show part of the syntactic structure (below the dotted line), but it gives no morphology and the semantics is incomplete. The main point is to show how apparent one-place predicates like "linguist", "yawning", "bald" and "loudly" can be shown as relationships between distinct concepts. (For more details see Hudson 1990: chapter 7 and 1995.)

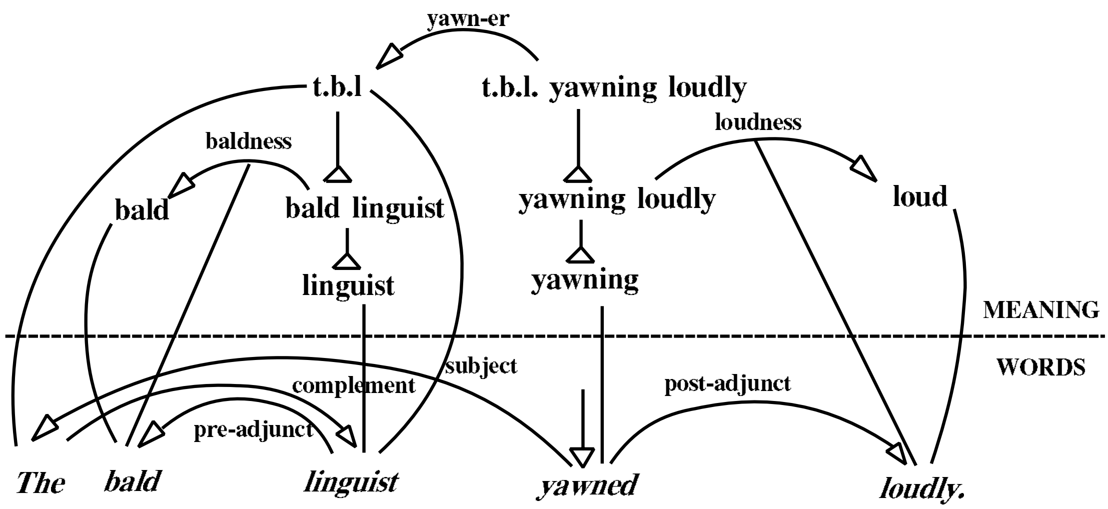

Figure 5

2.8 Conclusion

We have reached two general conclusions about the network view of language. First, contrary to what I myself once thought (Hudson 1990:15), it is not just a matter of notation, an alternative way of presenting information which could have been presented as a list of propositions. It is true that a network can be "unpacked" as a set of propositions, but the network model places specific constraints on propositions, such as the fact that they have to be binary. Furthermore, networks have properties that a set of propositions does not have, such as the possibility of measuring conceptual distances in terms of intervening links. Above all, only a network can serve as the database for a spreading-activation processor.

The second general conclusion is that there is considerable evidence to support the network model - evidence from experimentation (e.g. priming), from speech errors, from neuropathology and even from quantitative sociolinguistics. All these sources of evidence converge on a view of language which is very different from the more familiar "box" model, in which cognition contains a box of language which in turn contains distinct boxes of syntax, morphology, lexicon and so on. Instead, we can see language as part of the "sprawling network" of cognition, even though it may be separated from other parts by regions of relatively low-density connections; within this subnetwork we may be able to find further sub-subnetworks and so on, but every part of the total network is organised along the same lines (around isa hierarchies) and shares nodes with other subnetworks.

In the rest of the paper I shall spell out some consequences of this view for specific areas of language analysis.

3 Polysemy and homonymy

What light does the network view throw on the traditional distinction between polysemy and homonymy? This distinction is recognised as a major problem in lexical semantics, and indeed as a practical problem for lexicographers, because it requires a binary distinction where everyone agrees there is in fact a continuum. The distinction has to be made wherever we find a range of alternative meanings which can be expressed by a word of the same form, because we have to decide whether or not the meanings all belong to the same lexeme. According to the traditional view the two extremes of the continuum are occupied by polysemy (one lexeme, many meanings) and by homonymy (two lexemes, one form). A standard example of polysemy is *DOG*, whose sense may be either "male canine" or just "canine"; and a standard example of homonymy is $BANK_{money}$ and $BANK_{river}$, whose senses are unconnected. It is clear in these cases whether the senses are closely related (polysemy) or extremely distant (homonymy), but in between such examples lies a continuum of less clear cases: *DRY* (applied to wine or soil), *CYCLE* (bicycle or abstract repetition), *THROW* (stone or party), and so on.

The assumption that lies behind all these discussions is that one of the functions of a lexeme is to show semantic relatedness. By assigning two senses to the same lexeme, the analyst is guaranteeing that these senses are closely related. But as we saw earlier a network analysis of meaning already shows the degree of semantic relatedness in terms of "cognitive distance" which is a continuum. Presumably there is just one link (an isa link) between "canine" and "male canine", contrasting with perhaps ten links in the chain from "money-bank" to "river-bank", and some intermediate number for the chain from the dry-ness of wine to the opposite of wet. When the network already shows semantic relatedness so sensitively, it seems perverse to use the very blunt instrument of lexeme contrasts to impose a binary cut in the same information.

Once freed from this obligation to show semantic relatedness, the lexeme can be used for the job that it is uniquely suited to perform: that of expressing the coincidence of syntactic, morphological and phonological characteristics. More strongly, lexemes must never be redundant - they must always "do some work" in the grammar. For example, suppose we recognise two *BANK* lexemes; they would be grammatically identical - same morphology (plural: *banks*), same word class, same lack of syntactic valency. Therefore one of them must be redundant because its meaning could be assigned to the other without losing any information at all in the grammar; so they should be collapsed into a single lexeme. In contrast, the verb *BANK* (as in *He banked the money*) must be a distinct lexeme in spite of its close semantic links to the "money-bank" sense of the noun.

The result of this analysis is shown in Figure 6, where the division into lexemes follows the grammar and not the semantics. The crucial point about this diagram is that the meaning "money-bank" is much nearer, in terms of links, to the meaning "banking" than it is to the meaning "river-bank". The chain which links "money-bank" to "river-bank" is just a place-holder for a properly justified analysis, but it is hard to imagine a shorter chain between these two concepts; whereas the one from "banking" to "money-bank" is both short and well founded (banking means putting something into a money-bank). This analysis shows exactly the same relationships as a different analysis in which the two traditional *BANK* lexemes are distinguished, so this distinction must be redundant, and wrong.

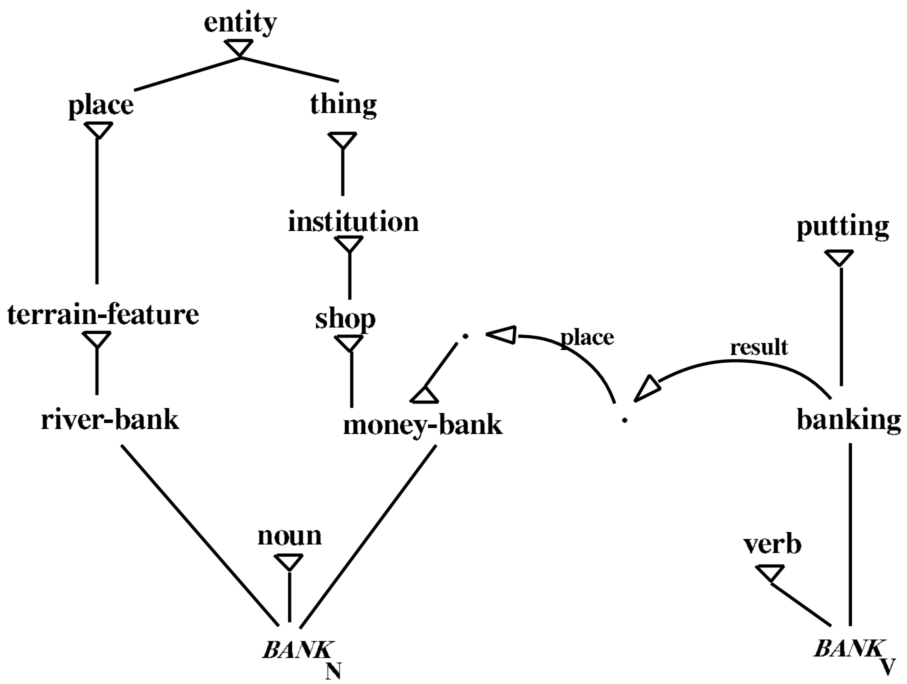

Figure 6

The conclusion, then, is that a network is a very good way of showing degrees of relatedness between words, whether this relatedness is a matter of sound, of syntax or of meaning. There is no point in trying to force a binary distinction between "similar" and "different" meanings in terms of one lexeme versus two.

4 Regular and irregular morphology

Another area where the network view is relevant is in the treatment of inflectional morphology (Creider and Hudson forthcoming, Hudson unpublished). The widely held "Dual mechanism" model makes a fundamental distinction between regular and irregular morphology on the grounds that regular morphology is handled by rule, while irregular morphology is expressed as default inheritance in an inheritance hierarchy (Marcus et al 1992). Such a view is incompatible with network analysis because there is no formal distinction between rules and default inheritance: "rules" are expressed as defaults which are inherited unless blocked. Depending on how we describe the information in a grammar, either everything is rules, or everything is network links; but there is no division into rules and links. The network view is therefore supported by the research which supports a model in which a single mechanism handles all morphology (Plunkett and Marchman 1993, Matcovich 1998). It remains to be seen exactly what kind of network model is needed - a connectionist model, or a more "symbolic" one such as I am proposing.

As we saw in Figure 1, WG morphology is based on two functions, "stem" and "whole". A word's stem is the form that it inherits from its lexeme, and its whole is its fully inflected form; and inflectional morphology handles all differences between the two, while derivational morphology deals with relations between the stems of different lexemes. (A language in which every word's form is that of its lexeme has no inflectional morphology, and needs no distinction between "stem" and "whole".) A WG morphology consists of a set of "equations" each of which defines a different function:

(5) a The whole of a $verb_{past}$ is its stem + its suffix.

b The suffix of a $verb_{past}$ is *ed*.

Equation (a) is diagrammed in Figure 1, where "$verb_{past}$" is the name of a sub-class of verb. If we assume that the past tense of *WALK* isa both *WALK* and $verb_{past}$, it follows by default inheritance that its whole must be *walked*. Diagram notation is particularly helpful when the right-hand side of the equation contains a functional definition.

For present purposes, the most important fact about WG morphology is that exceptions, general defaults and sub-generalisations all have the same formal properties. We can illustrate this point through a tentative analysis of a fragment of the Norwegian verb system, focussing on past-tense morphology (Simonsen 1998, Endresen 1998). Most forms in a regular verb paradigm have a suffix, so by default a verb consists of its stem followed by some suffix, and in a past-tense verb, this suffix is generally *et* (e.g. *KAST*, "throw", past *kastet*). This defines the so-called "Large weak" class which we can call simply "verb". Like English, Norwegian also has a class of "Strong" verbs, which have no suffix and generally have a different stem-vowel in the past (e.g. *DRIKK*, "drink", past *drakk*); for these we suppress the default suffix, and introduce another function, "stem-vowel" which maps onto the last vowel of the stem. (We must take this mapping for granted here.) The stem's default vowel is replaced in the past tense by another which varies from verb to verb (with sub-regularities that we shall ignore). This prose description shows how both regular and irregular verbs are analysed in terms of the same combination of functions and equations. These rules are diagrammed in Figure 7.

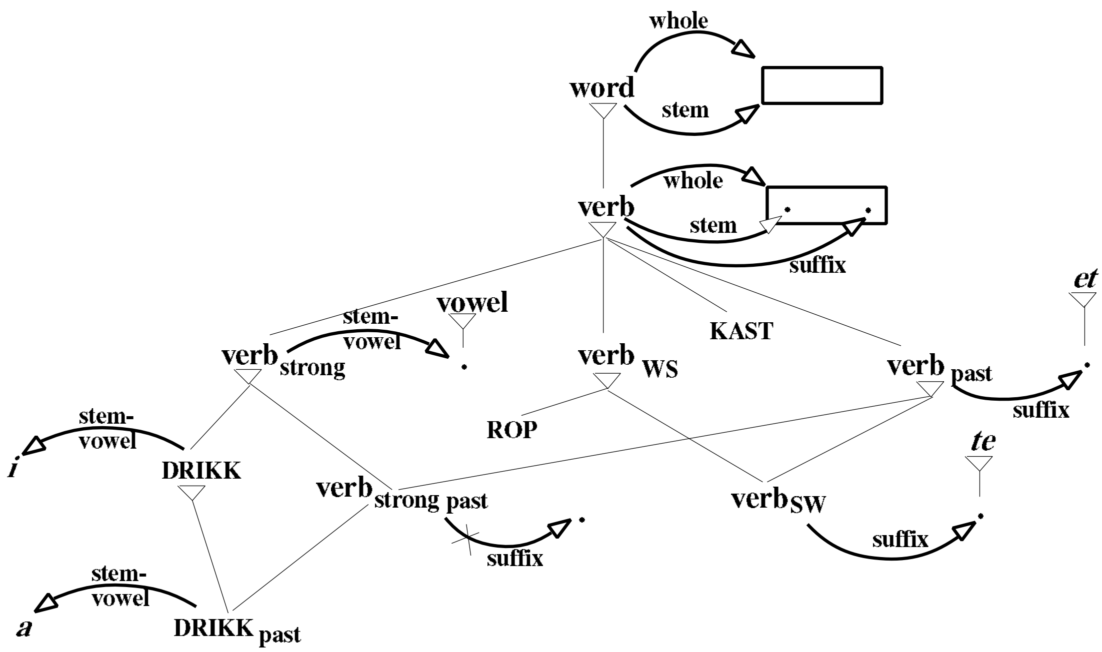

Figure 7

Unlike English, however, Norwegian also has an important subclass of weak verbs which have a number of peculiarities, one of which is that their past-tense suffix is *te* instead of the default *et* (e.g. *ROP*, "call", past *ropte*). These verbs are also shown in Figure 7 as the "Small weak" class (abbreviated here to SW). This sub-regularity is treated in just the same way as the regular and irregular patterns: we recognise a sub-class of verb and assign to it characteristics which override the defaults. Sub-regularities such as this are a serious problem for the dual-mechanism theory because the "rules" for them are themselves exceptions.

5 Conclusion: symbolic networks

The main purpose of this paper has been to argue that the whole of language (not just the lexicon) is a network with quite well-understood and consistent formal properties based on default inheritance and binary functions. This conclusion is completely compatible with the Cognitive-Linguistics assumption that language is part of the much larger network of general cognition.

The one issue which we have not considered is how these networks fit into the contrast that Pinker (among others) draws between "mentalese", the symbolic language of thought, and "connectoplasm", thought as defined by a connectionist version of a neural network (Pinker 1997:128). Pinker believes that the mind is divided between the two, though it may be unclear where the boundary lies: "Where do the rules and representations in mentalese leave off and the neural networks begin?" (ibid:112). The two systems have quite different formal properties: mentalese is symbolic, "digital", rule-governed, and uses concepts that are "crisply" defined, whereas connectoplasm is an "analog" network in which concepts are not defined (as such) at all, but related only by degrees of similarity induced from experience. Given these formal differences, it is unclear how the two parts communicate, so alternatives should still be considered.

One alternative is that it is "networks all the way down", as claimed in connectionism, but that at least some of the networks are much more like Pinker's mentalese. We can call them "**symbolic networks**". Symbolic networks are widely used in cognitive modelling, and indeed Pinker himself recognises the role of so-called "semantic networks" within mentalese (ibid:87) as a way of displaying a propositional database. The same is true of the networks of cognitive linguistics, including those that I have advocated in this paper. A symbolic network is symbolic, like mentalese, in that each concept has just one node and a finite and relatively fixed list of properties; but it is a network because the properties are defined by links to other concepts. It can be called the "language of thought" because it has the same formal characteristics as human language - namely, those of a network; thus it is the network that allows compositionality and recursion, which Pinker picks out as especially problematic for connectoplasm. As we have seen, it is merely a matter of terminology whether we call the links in a network "relations" or "rules", so we can say that a symbolic network is rule-governed like mentalese. In short, symbolic networks have all the properties that Pinker attributes to mentalese, so we may be able to conclude that mentalese is a symbolic network.

References

Bennett, David

1992 Stratificational Grammar. In Bright, W. (ed.) International Encyclopedia of Linguistics, Vol 4. Oxford: Oxford University Press, 77-9.

Chomsky, Noam
1986 Knowledge of Language. Its Nature, Origin and Use. New York: Praeger.

Collins, A. and Quillian, M.
1969 Retrieval time from semantic memory. Journal of Verbal Learning and Verbal Behavior 8, 240-7.

Creider, Chet and Hudson, Richard
forthcoming Inflectional morphology in Word Grammar. Lingua.

Emmory, Karen and Fromkin, Victoria
1988 The mental lexicon. In F. Newmeyer (ed.) Linguistics: The Cambridge Survey, Vol 3. Cambridge: Cambridge University Press, 124-49.

Endresen, Rolf Theil
1998 Problems in the morphological description of Norwegian verbs with *te/de* in the past tense. Paper presented at "The Verb in Cognitive Grammar" workshop, June 1998.

Fodor, Jerry
1983 The Modularity of Mind. An Essay on Faculty Psychology. Cambridge, MA: MIT Press.

Fromkin, Victoria
1988 Grammatical aspects of speech errors. In F. Newmeyer (ed.) Linguistics: The Cambridge Survey, Vol 2. Cambridge: Cambridge University Press, 117-38.

Goldberg, Adele
1995 Constructions. A Construction-Grammar Approach to Argument Structure. Chicago: University of Chicago Press.

Halliday, Michael
1985 An Introduction to Functional Grammar. London: Arnold.

Hudson, Richard
1971 English Complex Sentences. An Introduction to Systemic Grammar. Amsterdam: North Holland.
1976 Arguments for a Non-transformational Grammar. Chicago: University of Chicago Press.
1984 Word Grammar. Oxford: Blackwell.
1990 English Word Grammar. Oxford: Blackwell.
1995 Word Meaning. London: Routledge.
1996 Sociolinguistics, 2nd edition. Cambridge: Cambridge University Press.
1998 English Grammar. London: Routledge.

unpublished *I amn"t.*

Jackendoff, Ray
1996 The Architecture of the Language Faculty: A neominimalist perspective.

Lakoff, George
1987 Women, Fire and Dangerous Things. What Categories Reveal about the Mind. Chicago: University of Chicago Press.

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

Langacker, Ronald
1987 Foundations of Cognitive Grammar, I. Theoretical Prerequisites. Stanford: Stanford University Press.
1990 Concept, Image and Symbol. The Cognitive Basis of Grammar. Berlin: Mouton de Gruyter.
1994 Cognitive Grammar. In R. Asher (ed.) Encyclopedia of Language and Linguistics. Oxford: Pergamon, 590-3.

Luger, George and Stubblefield, William.
1993 Artificial Intelligence. Structures and Strategies for Complex Problem Solving, 2nd edition. New York: Benjamin/Cummings.

Marcus, G. F., Pinker, S., Ullman, M., Hollander, M., Rosen, T. J. and Xu, F.
1992 Overregularization in Language Acquisition. Monographs of the Society for Research in Child Development 57, No 228.

Marslen-Wilson, William
1980 Speech understanding as a psychological process. In Simon, J.C. (ed.) Spoken Language Generation and Understanding. Boston: Reidel

Marslen-Wilson, William and Tyler, Lorraine
1980 The temporal structure of spoken language understanding. Cognition 18, 1-71.
1981 Central processes in speech understanding. Philosophical Transactions of the Royal Society of London B295, 317-32.

Martin, James
1992 Systemic Grammar. In Bright, W. (ed.) International Encyclopedia of Linguistics, Vol 4, 120-2. Oxford: Oxford University Press.

Matcovich, Paola
1998 The Acquisition and Processing of Past Participle Morphology in Italian: An Experimental Study. University of Oslo Cand Philol. Thesis.

Pinker, Steven

1994 The Language Instinct. Harmondsworth, Middlesex: Penguin.
1997 How the Mind Works. Harmondsworth, Middlesex: Penguin

Plunkett, Kim and Marchman, V.
1993 From rote learning to system building: Acquiring verb morphology in children and connectionist nets. Cognition 48, 21-69.

Reisberg, Daniel
1997 Cognition. Exploring the Science of the Mind. New York: Norton.

Simonsen, Hanne Gram
1998 Past tense acquisition in Norwegian: experimental evidence. Paper presented at "The Verb in Cognitive Grammar" workshop, June 1998.

Swinney, D.
1979 Lexical access during sentence comprehension: (Re-)consideration of context effects. Journal of Verbal Learning and Verbal Behavior 5, 219-27.

Taylor, John
1989 Linguistic Categorisation: An Essay in Cognitive Linguistics. Oxford: Oxford University Press.

Yelland, G. W.
1994 Word recognition and lexical access. In R. Asher (ed.) Encyclopedia of Language and Linguistics, Oxford: Pergamon, 5009-14.

Endnote

1. The books about Word Grammar are Hudson 1984, 1990, 1995, 1996, 1998.

The following web site holds downloadable material, including a 30-page

introduction and a 100-page encyclopedia which is updated annually:

http://www.phon.ucl.ac.uk/home/dick/wg.htm