Word Grammar, cognitive linguistics and second-language learning and teaching

Richard Hudson

1 Linguistic theory and teaching

1.1 A personal introduction

To start on a personal note, this paper is my first opportunity to bring together two strands of my working life which have so far been quite separate (Hudson, 2002a): building models of language structure (Hudson, 1976, Hudson, 1984, Hudson, 1990), and building bridges between academic linguistics and school teaching (Hudson, 2004, Hudson & Walmsley, 2005). Of course, the two strands have influenced each other in my own thinking but I thought it would merely have confused both goals to combine them. After all, I was arguing that a cognitive model was right for language because it was true rather than because it was useful; whereas the arguments for applying linguistics in school teaching were all about usefulness rather than truth. However I was always aware that truth and utility are hard to separate: ultimately, the most useful approach to solving a problem must, surely, be one based on a true understanding of the problem, and ultimately, a true theory must apply to the real-world situations that we define as problems. I very much regret the historical trends of the last few decades which, at least in the UK, have pushed 'theoretical linguistics' and 'applied linguistics' further apart, and welcome the opportunities that cognitive linguistics provides for bringing them together again.

Another personal card that I should lay on the table now is that most of my school-oriented thinking and activity has been directed at teaching of the mother-tongue (L1) rather than of a second language (L2), and that this is also my first opportunity to write about the learning and teaching of L2 as such. However, I accept the argument of the 'Language Awareness' movement that L2 teaching should build on a general ‘awareness’ of how language works (Hawkins, 1999; Renou, 2001; Wray, 1994), so I see L1 and L2 teachers as contributing jointly to a single intellectual area of skill and understanding: ‘language’. Consequently, I believe that an L1 perspective is highly relevant to the topic of this book.

Because of these personal biases, my discussion of L2 learning and teaching will be oriented to the most typical language-learning situation in the UK, where a monoglot English child is learning a foreign language such as French in school. I am of course aware that there are many other L2-learning situations, and I hope that some of the general conclusions of this paper will extend to them as well.

1.2 Variation in language

L2 learning raises well-known questions for any theory of cognition and language, such as how bilinguals store and access their language, but there are other questions that have hardly been touched on. One such, which I have personally found particularly intriguing in the last few years, is the prowess of ‘hyperpolyglots’, people who know dozens of languages well; for example, the legendary Cardinal Giuseppe Mezzofanti was said to be familiar with 72 languages and fluent in 39 (Crystal, 1997:364, Erard, 2005). If they can do it, why can’t the rest of us? More interestingly still, there are communities where every single adult speaks several languages fluently; as far as I have been able to determine through informal email inquiries, the highest number of languages spoken in this way is six. (According to Hilaire Valiquette, this is true of Wirrimanu (Balgo), Western Australia.) Even if hyperpolyglots have very special brains, this cannot be true of whole communities so we all have brains which are capable of holding at least six entire languages. Why do most of us have such difficulty in learning just one or two foreign languages? And how do successful learners cope cognitively with such large numbers?

Such questions go to the heart of our ideas about the nature of language. If we have an innate language module, how can it accommodate more than one language? If we have innate parameters to set, how can bilingual children set them differently for several languages at the same time? What is the difference between a ‘mother tongue’ or ‘first language’ and a second language for a child who learns several different languages simultaneously? How do fluent bilinguals manage to speak in just one language when dealing with a monoglot, but to switch effortlessly in mid sentence when in bilingual company? These questions show the need for a model of language structure and language learning in which it is normal to learn and know more than one language. Most obviously, the model will need a place for the notion ‘language X’ (e.g. English, French and so on); but somewhat surprisingly, perhaps, very few theories of language structure satisfy even this modest demand. Even the most modern grammars appear to be torn between the traditional aim of describing ‘a language’ and the modern cognitive aim of describing a speaker; and the result is that we all pretend that these goals are the same. For multilingual speakers they are most certainly different.

Even for monolingual speakers this pretence turns out to be deeply problematic as soon as we consider the details of what typical speakers actually know and do. No speaker’s individual competence is neatly limited to what we might want to call a single language. On the one hand, many monoglots know fragments of other languages – loan words, quotations, even personal names – and on the other hand, every individual exhibits what Labov calls ‘inherent variability’ – different ways of saying the same thing which vary systematically with the social context, and which crucially vary within the speech of one individual. For example, a typical English speaker has two pronunciations for the suffix –*ing*, as in *walking*: with an alveolar [n] or a velar [ŋ] or [ŋg], called collectively the (ing) variable. The choice between these variants is free, but they are linked symbolically to different social groups and social situations so their relative frequencies can be related statistically to objective social variables (Labov, 1989).

Like a bilingual's choice between two languages, the (ing) variable offers a linguistic choice which is based on social choices and raises the same basic question for linguistic theory: how to accommodate linguistic alternatives in a single individual language system without ignoring relevant social distinctions. Unfortunately, we are still waiting for a fully worked-out answer, but I have made a start (Hudson, 1996:243-257, Hudson, 1997, Hudson, 2007) and although the hard work remains to be done, I believe that the basic ideas are relevant and important for the teaching of L1 and L2.

In short, I believe linguistic theory and description are ready to emerge from the isolation that structuralism demanded at the beginning of the twentieth century. At that time it was important for linguists to concentrate on the structure of the language system, ignoring external factors such as the supposed world-view or level of civilization of the speakers. But by the end of the century the focus had shifted from the language system to the individual speaker's cognitive system – a very different object, especially if we consider multilingual speakers. The next step will therefore be to develop theories of language with L2 and internal variation at their centres. Most importantly of all, perhaps, they will have a place for linguistic attitudes – the feelings that different languages and variants evoke. We have a great deal of empirical data about these attitudes, as well as a framework of ideas from social psychology for interpreting them (Giles & Powesland, 1975, Giles & Bradac, 1994), but we are still some way from being able to integrate them into a cognitive model with the rest of language structure, though I shall suggest below that we can already take some steps in this direction. (The same problem arises with emotional language such as swear-words and exclamations, whose emotional content has no place at all in any existing model of language.) This is an unfortunate gap for L2 research, where linguistic attitudes are central, so I hope it will receive more attention in the cognitive-linguistics community than it has so far.

The cognitive enterprise is an ambitious one because any cognitive theory of language has to cover not only language structure, but also how it is learned and used and how speakers feel about it. I shall suggest that cognitive linguistics in general, and **Word Grammar (WG)** in particular, are well placed to at least start meeting this challenge, but I am well aware of the work that remains to be done before it is a reality.

2 Language is a network

2.1 Vocabulary is a network

One of the leading ideas of cognitive linguistics is that language is a cognitive network of units – meanings, words, sounds and so on (Hudson, 1984; Goldberg, 1995; Langacker, 2000). Indeed, it could be argued that this is also the basic idea behind all structuralism, with its emphasis on interconnections and systems; but it goes beyond mere structuralism because it denies that there is anything else in language. In particular, there are no 'rules' as such, although (of course) there are plenty of generalizations. I personally first met the idea that language is nothing but a network in the work of Lamb (Lamb, 1966; Lamb, 1998), and I now think of it as the main claim of WG (Hudson, 1984:1, Hudson, 2000; Hudson, 2007). I also think it is a particularly important idea for L2 researchers, as I shall explain below.

The network idea is particularly obvious when applied to vocabulary. A typical word is at the centre of a network of homonyms, synonyms, word classes and collocates, each of which is at the centre of another little network. The network for just one word, the adjective FAST, is shown in Figure 1, where the little triangle is standard WG notation for the 'isa' relation between a concept and its super-category (e.g. BIG 'isa' adjective). I leave it to the reader to imagine this network when expanded by the extra links for the verb FAST, the adjectives BIG and QUICK, and the collocating noun WORKER (as in *fast worker*). But even this expanded network is tiny compared with the total network for any one person's knowledge of English.

[Insert Figure 1 about here]

The research evidence for this network structure is rock solid and virtually beyond dispute. It is based on the fact that when we use some concept, we make the node for that concept active by focusing energy on it; but crucially, this process is rather messy as energy spills over onto the neighbouring nodes as well. This is called 'spreading activation'. Take the tiny network in Figure 1 for example, where the concepts include words; if we either say or hear the adjective *fast*, then we must be using the concept labeled $FAST_A$ so this node becomes very active, but all the nodes that are directly connected to it also become more active than they would otherwise be, thanks to the blind spreading of activation. There are two kinds of research evidence for this claim, involving speech errors and 'priming'. When we accidentally use the wrong word, the word that we use is almost always a 'network neighbour' of the target word; for example, when someone used the word *orgasms* when they intended to use *organisms*, the mistake presumably happened because the two words are so close in terms of pronunciation (Aitchison, 1994:20, Harley, 1995:360). This preference for network neighbours obviously presupposes a network structure within which activation can spill over from one word (or other concept) to its neighbours. Similarly for priming, where timing experiments have repeatedly shown that a word is easier to find – for example, in order to decide whether or not it is an English word – if it has just been preceded by a network neighbour; for example, it takes less time to decide that *doctor* is an English word if one has just heard the word *nurse* than after hearing an unrelated word such as *lorry* (Harley, 1995:17).

2.2 The vocabulary network of L2

For the L2 teacher or researcher, the main point of this discussion of networks is that the target – L2 – is a network just like this, and not a list of disconnected items. When you learn a new item of vocabulary you add a little network of links to other nodes which define its pronunciation, its meaning and so on by linking it to items in the existing network. Moreover, you cannot use the item on future occasions unless you can find it, which you do by activating it via the given clues – via its pronunciation when listening and via its meaning when speaking. Since this activation reaches it by spreading from neighbouring nodes, the more links it has, the easier it is to find; and if you cannot find it, you have essentially forgotten it. A word (or any other concept) which only has a couple of links is barely integrated and easily forgotten, but a rich collection of links guarantees the word a long and useful life. (I expand on this point in 3.1.)

Clearly the L2 network will be small and fragmentary to start with, but the L1 network is vast and rich, so the obvious strategy is to build the network for L2 onto the existing one for L1. A translation equivalent provides one such link ("*chat* means 'cat'"), but other anchor-points come from form-based links (e.g. *chat* and *cat* have the same spelling except for one letter). Mnemonics provide further links for the beginner at the point where the main object is simply to make the new word into a permanent part of the network rather than to build links which will be useful in future processing (e.g. *chat* contains the 'hat' of the Cat in the Hat). Any link is better than no link (Meara, 1998). Figure 2 is meant to model the effect, on an English speaker's language network, of learning the French word CHAT (though I have deliberately simplified the network in some respects, and omitted any mnemonic links). The solid lines are links that already existed for the English word CAT, plus the category 'French word' which presumably exists for someone who has already started to learn French; in contrast, the dotted links are learned.

[Insert Figure 2 about here]

This network shows how intimately connected the two languages must be in the learner's mind; in other words, I am assuming a 'common-store' model of bilingualism (Harley, 1995:133). Though this is controversial among psychologists, I can see no alternative in a network model because the learner so obviously builds on pre-existing knowledge such as the letters of the alphabet and the meaning. There is also empirical evidence that the word classes are shared even in fluent bilinguals who switch languages in mid-sentence. In such cases, the word classes of the two classes are virtually always respected – a phenomenon known as ‘categorial equivalence’ (Muysken, 2000); for example, if an English verb requires a noun as its object, then whatever the language of the object, it will be a noun. This constraint only makes sense if the two languages share the same word classes so that the language contrast cuts across the word-class divisions as in Figure 2. Moreover, if different languages share the same word classes even in the mind of a fluent bilingual, how much more likely it is that the same will be true in the mind of a beginner who knows nothing about word classes in the new language.

We can safely assume, therefore, that words from different languages will share the same set of word classes in the mind of the learner – except, of course, where the languages have different structures which require different word classes; for example, only one language may have classifiers. However, another kind of category that the two languages must share is their relations. As the diagrams have already shown, the kind of cognitive network that I am describing is not merely a set of associations, but a highly articulated structure where every link belongs to a specified type. In particular, I am advocating what is called a ‘local’ or ‘symbolic’ network, in which every concept corresponds either to a node or to a link, and not a ‘distributed’ or ‘connectionist’ network in which nodes and links are undifferentiated and the distinctive content consists entirely of activity settings. The ‘isa’ relation is fundamentally different from all the others, but the others also fall into categories such as ‘meaning’, ‘pronunciation’, ‘spelling’ and ‘dependent’. (The same is true for all flavours of cognitive linguistics, but I think WG has a particularly well developed theory of relations.) Once again, it is hard to imagine any alternative to the assumption that L2 builds on the existing categories for L1; so in Figure 2 the same relations ‘meaning’ and ‘spelling’ are found in the French word as in the English one.

2.3 Grammar is a network too

The discussion so far has revolved around vocabulary, where a network analysis is more or less uncontroversial; a moment’s serious thought about the relations among words is enough to show that there is no serious alternative (Aitchison, 1997, Pinker, 1998). It is much less obvious that, or how, the same kind of analysis can be applied to the more general patterns of morphology, syntax and phonology. However, it is common ground among cognitive linguists that it can. Moreover, the same kind of empirical support is available as for networks in vocabulary. Here too, we find speech errors that can only be explained in terms of spreading activation in a network. For example, when someone said *slicely thinned* (Levelt, Roelofs, & Meyer, 1999), they must have activated the general concepts ‘adverb’ and ‘past participle’, together with their default morphology, separately from the lexical items SLICE and THIN. A similar explanation applies to *I’m making the kettle on,* a blend of *I’m making a cup of tea* and *I’m putting the kettle on* (Harley, 1995:355); in this case, both PUT and MAKE were candidates for expressing the intended meaning, and the general syntactic pattern of direct object followed by ON must have been activated along with PUT so that it could then be combined with the other verb. Moreover, priming experiments also show that general patterns can be activated. For example, *Vlad brought a book to Boris* primes other sentences with the syntactic pattern consisting of a direct object and prepositional phrase (Harley, 1995:356, Bock & Griffin, 2000).

All this evidence supports the idea that general patterns are stored and processed in much the same way as more specific patterns.

The conclusion, therefore, must be that cognitive networks contain general concepts as well as specific ones, and that there is some kind of logic which allows us to connect the two – in other words, which allows us to apply generalizations. Cognitive linguists agree in recognizing schematic concepts such as ‘Word’ or ‘Animal’ which carry general properties and are linked to more specific concepts by a special relation which is dedicated to this role. In WG this is the ‘**isa**’ relation which I have already introduced; for example, CAT isa Noun, CHAT isa French-word. This relation carries the basic logic of generalization called ‘**inheritance**’, whereby sub-concepts ‘inherit’ properties from their super-categories; for example, if CAT isa Noun, and Noun has a number, then CAT must have a number as well, and if CHAT isa French-word and Noun, and we know that French nouns have a gender, the same must be true of CHAT.

Inheritance is a very powerful mechanism for handling any kind of generalization, and in language it replaces the notion of a ‘rule’. For instance, instead of saying ‘A verb takes a noun as its subject before it’ we define the relation ‘subject’ between Verb and some word W which combines various properties:

being a noun (W isa Noun)

being obligatory (the quantity of ‘subject’ is 1)

preceding Verb

Any word which isa Verb automatically inherits this relation and all its properties as shown in Figure 3, where the dotted properties of *loves* are inherited from the super-category Verb. Incidentally, syntax is particularly simple in WG because it is based on relations between single words; phrases turn out to be redundant when all syntactic relations are made explicit (Hudson, 1990:105, Hudson, 2007).

[Insert Figure 3 about here]

The same kind of network translation is possible for morphological rules, although these are usually expressed as processes, e.g. “To form the plural of a noun, take its stem and add the suffix {s}.” Once again, we need a general word class (Plural) but this time we also need to invoke an extra level of ‘forms’ which lies between words and sounds. Words are related to forms by the relations ‘stem’ and ‘word-form’ (i.e. the fully inflected form), and special ‘variants’ of forms are defined by relations such as ‘s-variant’. A simple analysis for plural nouns would contain the following:

- A plural noun has a stem S.
- It also has a word-form F.
- F is the s-variant of S.

The ‘s-variant’ relation is a step forward because it is purely morphological, so it can be re-cycled for other syntactic word classes such as singular verbs – in other words, it helps with systematic syncretism. The next step relates s-variants to the morpheme {s}:

- A form has an s-variant V.
- V has two parts, Part1 (A) and Part2 (B).
- A is a copy of the form.
- B isa the morpheme {s}.

These two patterns of generalization are shown in Figure 4, so they are available for inheritance by any plural noun thanks to its isa link to Plural.

[Insert Figure 4 about here]

The logic of inheritance goes much further than the simple copying of properties because it only copies 'by default' – hence the technical name '**default inheritance**'. This is the logic of common-sense reasoning which allows exceptions such as three-legged cats and unusually large birds, and in language it is the logic of every descriptive grammar which recognizes generalizations and exceptions. The exceptions that come to mind first are probably irregular inflections (e.g. *geese* as the plural of *goose*, blocking the default *gooses*), but these are actually just the tip of a very large iceberg which includes most of the complications of syntax such as non-default word order (Hudson, 2003). For example, Figure 3 shows the default subject-verb order, but subject-auxiliary inversion provides an alternative to the default (as in *Are you ready?*). To handle this, we introduce a special 'inverting' sub-class of auxiliary verbs with special verb-subject order (as well as the semantics of a question); the analysis is shown in Figure 5. Like a word-order transformation, this captures the important insight that *You are* is 'underlying', 'unmarked' or 'basic' in comparison with *Are you*.

[Insert Figure 5 about here]

The big generalization that emerges from this discussion is common ground to all cognitive linguists: 'the lexicon' (containing vocabulary) and 'the grammar' (containing general rules) are just two areas of a single network. The lexicon deals with relatively specific word-types such as CAT, whereas the grammar deals with relatively general categories such as Noun, but both areas have the same network structure and there is no boundary between them. This is probably the point where cognitive linguists part company most clearly from the earlier generative tradition, with its rules and lexical items, and recent psycholinguistic discussions of morphology have tended to contrast unified approaches such as this with a 'dual-mechanism' analysis (Pinker, 1998). The debate is important for L2 research because it involves the fundamental nature of what is learned. If vocabulary and rules really are different, they could be learned in quite different ways, but if they are fundamentally the same we can expect a single mechanism for learning.

2.4 Language and memory

It is not only language that is a network: cognitive psychologists generally agree that the same is true of general long-term memory (Reisberg, 1997:257). How are these two networks related in our minds? Linguists are divided on this question, but cognitive linguists are agreed that the language network is simply the part of long-term memory that deals with language. Given the central role of words in WG, we can define language simply as our memory for words. In contrast, many linguists believe that language is a 'mental module' (or a collection of modules) which is clearly separate from the rest of knowledge both in terms of how the knowledge is stored and in terms of how it is used (Fodor, 1983; Chomsky, 1986; Smith, 1999). However, there are good reasons for rejecting the modular view (Hudson, 2002b; Hudson, 2007), so I shall assume that 'knowledge of language is knowledge' (Goldberg, 1995:5). This is not, of course, to deny the variety of knowledge – we still have to distinguish factual knowledge from perceptual knowledge (e.g. images, sounds, smells), motor skills and feelings. But in this classification, language is part of the same gigantic mental network that we use for our social and physical world, and which is often called 'long-term memory'.

Another traditional boundary separates long-term and short-term memory. The latter is often called 'working memory', in recognition of its role in exploiting the contents of long-term memory. One version of modularity is the idea that we have a dedicated working memory for language or even for specific areas of language such as syntax (Lewis, 1996). In this view, working memory is a separate area of the mind, like a workbench, onto which information is copied from long-term memory or perception and then transformed into some kind of output. However many psychologists now agree that working memory is just a convenient term for the activity – such as spreading activation and inheritance – which is known to take place in long-term memory itself (Miyake & Shah, 1999). This view is much more in harmony with cognitive linguistics with its emphasis on the unity of language and the rest of thought.

2.5 Separating different languages in a network

The final question about the formal structures involved in L2 learning is how the two languages are kept separate. As I mentioned earlier, most theories of language are designed to model monolingual speakers who know only one dialect and use only one style. This limitation is left over from the old days when linguistic theories applied to languages rather than to speakers, but is not part of a truly cognitive approach. Indeed, in WG there are several different mechanisms for including variation within the same network analysis.

I have already suggested one mechanism, exemplified by ‘French word’ in Figure 2. This is what we need for any new generalization about all French words, such as how to relate spellings to pronunciations, which are clearly part of what needs to be learned. For example, grammatical properties of French nouns will be attached to the category ‘French noun’, which isa both Noun and French word. No doubt those (like me) who learned just one language in early childhood treat this language as the default, so (in my mind) Word and English word are probably the same concept.

Another mechanism involves the notion of ‘a language’ which includes named languages such as English or French, each with whatever properties we may know about it – its name, where it is spoken, and so on. Clearly, languages are part of our knowledge of the world, so there is nothing controversial about this suggestion, but knowing about a language is obviously different from knowing that language and indeed from knowing anything at all about its words. However, if we do know some words of language L, we must have a way of relating each of these words to L, so we have a relation ‘language’ between words and their language, e.g. the language of CHAT is French. As an example, Figure 6 models my knowledge of at least three languages, and shows that I know absolutely no Urdu, just one word of Russian and enough French words to be able to generalize about them.

[Insert Figure 6 about here]

A third mechanism for handling variability in a person’s knowledge of language is the one which is needed for the inherent variability discussed in 1.2, where individual items of pronunciation or vocabulary are limited to particular kinds of speaker or situation; for instance, in my experience the word BONNY is limited to

Scottish speakers. This mechanism requires a link between a stored word or other pattern and a particular kind of speaker or situation. Situations are still hard to formalize, but speakers are easy because they are already needed for deictic words such as ME, whose referent is the speaker (Hudson, 1990:128). Examples like these make a rather obvious point unavoidable: that the language structures which we store in our minds are based on our experiences with language, so they may well include details of the situation such as the kind of person who speaks it. The two networks in Figure 7 illustrate these two important applications of the 'speaker' relation, but they also make the important point that because any word is primarily an action – an utterance – rather than marks on a page, it inherits an actor from the general 'action' category; and of course the 'actor' of a word is its speaker.

[Insert Figure 7 about here]

The 'speaker' link can apply to any portion of language whose speakers we can identify, from individual words to entire languages; for example, if we know that all speakers of Gaelic are Scots, then this is part of the knowledge structure for Gaelic. These links bring together two rich networks of knowledge: what we know about words and what we know about people – our linguistic and social networks. As with language, the social concepts vary in generality from individual people through various categories of people to the most general category, Person (compare: Word); and each concept brings together a collection of properties. These properties are all based, in some way, on experience, but this does not guarantee truth or fairness; on the contrary, social concepts are where we find prejudice and other kinds of social stereotyping, so thanks to the 'speaker' links these stereotypes intrude into language.

As we all know, the way a person speaks sets up all sorts of expectations about their personality and behaviour, and common experience is amply confirmed by the research on attitudes mentioned in 1.2. This research is important for L2 learning because attitudes attach as easily to whole languages as to variants of a single language.

3 How to learn a language network

3.1 From token to type

It should be clear by now that I believe that language is learned, just like any other kind of factual knowledge, rather than 'acquired' by the triggering of innate concepts. As I mentioned earlier, although this view is controversial in linguistics, it is more widely accepted by psychologists (whose opinion must surely count for more in this area). In any case, although as a linguist I am of course aware of the standard arguments in favour of innate concepts (e.g. poverty of the stimulus, selective impairments and critical periods), I find them all quite unconvincing (Hudson, 2007). For lack of space to defend my view I shall simply take it for granted. The main point for L2 is that it can be viewed as a body of knowledge like any other, to be learned and taught by experience. To the extent that it is distinctive, that is just because of its unique place in human life – a matter of function rather than form.

The first step in understanding how we learn a language network is to ask how we use it. How do we apply it to particular experiences, such as hearing and understanding a word such as *cat* in the sentence *Your cat is in our garden*, spoken by a neighbour? To make an obvious point, this uttered word is distinct from the stored word CAT, so they are held mentally as two distinct concepts. In technical terms, the (uttered) '**token**' is distinct from the (stored) '**type**'. It is all too easy to be misled by the conventions of writing (which do not distinguish types and tokens) into ignoring the difference, but it is actually rather important. For example, in perception we start with a mental representation of the token and its properties, and it takes some processing to work out which type it is an example of. Moreover, we can still link the token to some type even if its properties are slightly different from that type's; for example, we can cope with spelling mistakes, foreign accents and so on, so we can see the spelling *yelow*, and we assign it confidently to the type YELLOW without even noticing the misspelling. In this case we choose the word with the nearest spelling to the token, but this need not be so: we may trade a greater deviance of spelling for the sake of a closer fit in terms of syntax and meaning (e.g. when we read *there* as *their*, or vice versa). The general principle that we seem to follow is to adopt the type which provides the best global fit with the token and the current global scene – what is often referred to as the '**best-fit principle**'. This apparently powerful and mysterious process may be quite easy to model in terms of spreading activation if we assume that the most active candidate type (e.g. in this case, the most active word type) is the one to choose (Hudson, 2007). We shall return to the best-fit principle and its consequences below.

On the other hand, although types and tokens are distinct, they are in some sense 'the same'. This is precisely the relation covered by the 'isa' relation that we find among types (2.1): the token *cat* isa the type CAT, which isa Noun, and so on. Furthermore, this is the relation that carries default inheritance so it allows the token to inherit all the stored properties of the type (and of its super-categories right up to the top of the hierarchy) – a wonderfully effective way of applying stored knowledge to new experiences, which allows us to go beyond mere observables. For example, we hear the sounds for *cat*, and inherit its word class and its meaning, just as when we see an unfamiliar cat and inherit the fact that it will enjoy having its tummy stroked. In other words, tokens form a constantly changing 'fringe' attached (by 'isa' links) to the edge of the permanent network.

Now suppose we consider a token not as a processing experience but as a potential learning experience. Most tokens (of words or of any other kind of experience) are unremarkable, and we may assume that their life in the network is very short indeed – just a second or two before they decay and effectively disappear from memory. This is where the well-known limits of 'working memory' (2.4) apply, so we are doing well if we can remember more than nine or ten words in a sentence that we have just heard. If working memory is merely the currently active part of the network, this limitation must follow from the amount of available activation energy; so we deactivate past tokens as soon as we can in order to make more activation available for future use. But some tokens are so remarkable that we can recall them days or even years later, which shows that this deactivation is not automatic. One plausible explanation for the prolonged life of these tokens is that they were 'hyper-active' on first encounter, and one possible reason why extra activation might have been needed is that the usual processes for finding a suitable type ran into trouble. This would be the case if we heard an unfamiliar word, or a familiar word used in an unfamiliar way; in either case there is no existing type, so we have to work out unobservable properties instead of merely inheriting them. In short, we have to guess

– a very expensive process in terms of activation. And one effect of all this activation may be that the token node stays active so long that it never disappears.

My suggestion, then, is that hyper-active tokens turn into types; in other words, they become part of the permanent network. This process of **storage** is a combination of remembering and forgetting: remembering the interesting and hyper-active bits, and forgetting uninteresting details such as the time, place and speaker (though any of these may be sufficiently remarkable to be remembered). This is another way of saying that we learn the language system from usage, which is how cognitive linguists generally see language learning (Barlow & Kemmer, 2000; Bybee, 1999; Bybee & Hopper, 2001; Ellis, 2002; Langacker, 2000; Tomasello, 2003). However, it is important to stress that storage can apply to a token even if the best-fit principle has produced an established type for it, provided its properties are sufficiently noteworthy. (For example, the misspelling of *yelow* could produce a stored sub-type linked to the person who wrote it as a mistake to be held against them in the future.) This means that a given word-type (such as CAT) may end up with numerous particular instances stored beneath it as sub-types, together with whatever properties made them stand out. Each stored sub-type reinforces the type in the network by providing extra links, and if a number of sub-types share some property, this may be added to the type's property by the process of induction discussed in 3.2.

This model of storage has two potentially important consequences for L2 learning. The first concerns **frequency**. According to this model, we only store tokens with unusual properties – i.e. properties that cannot be inherited from the type. Familiar words already have a rich type that covers all their properties, so they are too boring to remember; but less common word types have had less chance to be enriched by the process outlined above, so their tokens are more likely to be noteworthy because of unpredicted properties. Consequently, there should be a much greater effect of frequency for rare words than for common ones – exactly as predicted by the power law of practice (Ellis, 2002). In the L2 context, hearing a rare word ten times in a lesson should produce much more effect on the learner's knowledge than the same number of repetitions of a frequent word. Moreover, we all know that memories for things such as names and new words are in danger of fading during the following days or weeks (in contrast with the seconds for which a token may survive). For evidence, we only need to consider students who cannot remember technical terms from one class to the next; but psychological studies also confirm that facts are at the greatest risk of being forgotten soon after being learned (Reisberg, 1997:246). The motto for the L2 learner is familiar: Use it or lose it.

The second consequence concerns **distinctiveness**. According to this model, a new type is strengthened by the addition of stored sub-types, but this will only happen if the best-fit principle can identify it as the best-fitting type for future tokens. A stored type that never wins the best-fit competition might as well not exist because it can never be retrieved, and it has no prospects of becoming more useful. But if the best fit is decided by spreading activation, a node's chance of success depends on how distinctive it is (along with other factors such as its base level of activation). A new word *zax* should be more memorable than *saf,* as every advertiser knows, provided it has the same number of subsequent repetitions. This is bad news for L2 learners, because one L2 word tends to look to the learner much like another one, so few words are sufficiently distinctive to be memorable in their own right. Fortunately this is not the end of the story, as there are ways to help a new word to survive even without future tokens; for example, it is possible to explore the word's lexical links to other words. But distinctiveness is clearly an issue for L2 learning and teaching.

3.2 Induction

Another part of learning a network is the induction of generalizations from these low-level stored types. This is a complex process because generalizations feed each other, and a false generalization may lead to a dead end. However, in principle the process is fairly simple, though there are many details that I (at least) don't know.

Generalizations in a network consist of correlated properties, i.e. properties that tend to co-occur. For example, the properties of flying, having a beak, building a nest and having feathers are strongly correlated in the sense that many different concepts have these properties and few of the properties occur without all the others; so collectively they justify the general concept 'bird'. Similarly, in language, the properties of combining with *the* or *a*, having an inflection with {s} and referring to a concrete object are correlated and justify 'noun'. In a network, properties are defined by links to other nodes, so nodes that 'share a property' must all have the same relation to some other node, and 'correlated properties' are distinct relations that link the same set of nodes to two or more other nodes.

An abstract set of correlated properties is shown in Figure 8. The three arrow-styles represent three distinct relations, each pointing at a different concept on the right but each linking this concept to almost the same set of concepts on the left. This congruence justifies the generalization that all three properties apply to A, B, C and D. Admittedly, the generalization is actually not quite true because concept A has a dotted link to W instead of Y, so this concept is an exception; but exceptions can be tolerated thanks to default inheritance.

[Insert Figure 8 about here]

According to the WG theory of learning (Hudson, 2007), whenever 'we' (i.e. our minds, but well below any kind of consciousness) notice a set of nodes with correlated properties, we record the fact by introducing a new node into the network, giving it the correlated properties, and giving it an isa link to each member of the set of existing nodes. This procedure turns Figure 8 into Figure 9, which now contains the generalization applied to a new node N, though all the original links are still there as well. This is how we manage to combine so many generalizations with so much fine detail: always build generalizations on existing fine detail, and never forget the fine detail.

[Insert Figure 9 about here]

How do we spot correlated properties? An honest answer is that I don't know, but once again I can speculate, and once again my answer has to involve spreading activation. This time I believe we have to think of rather low-level activation circulating through our memory in a rather random and undirected way – perhaps during the 'down time' discussed above when activation maintains concepts (and which sometimes wakes us in the middle of the night with a name or a word that we failed to find during the day).

I guess that this background activation must also be responsible for inductive generalizations. Take the abstract example in Figure 8. Suppose just one of the right-hand nodes X, Y or Z becomes active: nothing important happens and the activation just moves elsewhere. But if two of them happen to become active at the same time, activation spreads from both of them to the same set of nodes A-D, signaling a correlation. On the principle of giving more to those who already have most, these nodes receive all the available activation and spread it out to any other properties they may have, thereby increasing the significance of the correlation and possibly collecting even more activation for the sharing nodes. Once these nodes have reached some threshold level of activation they spawn a super-category with isa links to them and copies of their shared properties; and the induction is complete.

If this theory is right, it has important implications for theories of L2 learning. First, it presupposes large amounts of initial raw data as the basis for induced generalizations. Second, it presupposes significant amounts of down time for 'digesting' these data and finding correlations. But of course in the school context, down time is scarce and has to be shared not only with other subjects but also with the multiple demands of the emotional and cognitive life of a child. This may work well for L1 and for learning outside school, but it is unlikely to lead to much spontaneous induction in L2.

3.3 Motivation and attention

The need for motivation is self-evident, but WG theory not only helps to explain why this is so, but also why motivation may be in short supply. Once again the mediating variable is activation and the key idea is rather obvious: motivation decides where we channel activation. As I have just suggested, this applies to the background activation which reveals generalizations, but it applies even more obviously on the time-scale of working (alias short-term) memory, which I discussed in 2.4, where the relevant variable is attention. We are all surrounded by competing calls on our attention, but this is even more true of school-age language learners in a typical classroom situation; so it is hardly surprising that many teachers believe that trying to teach unmotivated students is simply a waste of time.

According to WG, then, motivation decides how we divide attention between competing claims; and what we mean by attention is simply activation. Activation requires some kind of energy, and since we only have a limited amount of this energy, activation is limited – hence the well-known limits on working memory mentioned earlier. This limited activation is a vital constraint on learning because there are only two ways to learn a new concept (short of direct instruction, to which we turn below): by turning a hyper-activated token node into a permanent type (3.1) or by inducing a general node to carry correlated properties revealed by background activation (3.2). Both of these processes require extra activation, so they both presuppose some degree of motivation.

Where does this motivation come from, and why might it not be enough for L2 learning? This is clearly a crucial question for the whole enterprise of L2 learning and teaching, and especially at school where children are by and large studying languages because they have to rather than out of choice. Of course, the same question arises for every subject in the curriculum, but languages are special because of the 'speaker' links which bind each language to the social stereotype of its speakers (2.5). If we consider the classic distinction between integrative and transactional motivation (Gardner & Lambert, 1972), they require either a very positive social stereotype or the prospect of practical benefits in the near future, neither of which can be taken for granted at all for most school-level language learners. At present the UK is facing a serious crisis in L2 study as decreasing numbers of school children choose to study languages at high school and university and language teachers become harder to recruit (Anon, 2002), and I believe that one of the reasons is that educationalists use transactional arguments (Kelly & Jones, 2003). Fortunately, there is a third motivation to consider alongside the classic pair: **interest**. I suspect that this is ultimately the only motivation that will drive English-speaking children to study foreign languages seriously at school, though adults are as open as any to integrative and transactional motivations. How to inspire interest is a topic I shall take up in the last section.

4 How to teach a language network

4.1 General considerations and the implicit/explicit contrast

Suppose cognitive linguists are right about language being a network learned from usage and firmly embedded in general knowledge. What implications should this have for the teaching of L2?

One obvious conclusion is that the school context for L2 is so different from the home context for L1 that we cannot leave anything to Mother Nature. A small child starts from scratch but has vast amounts of richly structured and helpful usage to learn from, and plenty of time to digest it. In contrast, a student at school starts with a richly structured language network for L1 but has much less experience of usage and even less time for digestion. Moreover, a child has devoted and intimate caregivers whereas a student has an expert teacher – a virtual stranger. The two learning situations could hardly be more different.

On the other hand, if language is just knowledge, then language teaching shares at least some features with the teaching of other kinds of knowledge – geography, maths and so on. As in these other subjects, the teacher starts with an explicit syllabus – a planned route through the network – and a repertoire of activities such as exercises to support the learning. There is a very clear content to be learned – the vocabulary, grammar and pronunciation of L2; and the content has a clear structure. Indeed, the structure of L2 may be easier to understand than that of other subjects precisely because linguists know so much about it. (How many other subjects have a content whose inherent structure is as well understood as language?)

It is true that L2 combines skills – speaking, listening and so on – with factual knowledge, but the same is true of maths, which is tested in terms of problem-solving abilities much like those of L2 practice. It is also true, as I said above, that L2 faces the learner with a social stereotype and its associated attitudes, which can affect motivation; but the same is true in subjects such as drama and religion. And it is true that L2 is not just a body of knowledge but is itself a tool for communicating knowledge; but the same is obviously true for L1, which the students are still studying alongside L2. In short, a cognitive approach reveals similarities rather than differences between L2 teaching and other school subjects. Indeed, we could even extend the idea of network structure to the school and point to the need for the L2 teacher to 'network' with teachers of other subjects.

This view of L2 has important consequence for the choice between implicit and explicit teaching. If the teacher wants students to learn some generalization, is it best to leave them to induce it from usage in which it is implicit, or to tell them the generalization explicitly? If L2 teaching is like other subjects, the question hardly arises. Even L1 teaching (at least in the UK) now favours explicit teaching at all levels (Anon, 1998), and it is self-evidently part of other subjects. Although it was out of favour in L2 teaching for some time it is now much more widely accepted (Norris & Ortega, 2000, Ellis, 2002; Ellis, 1994, Anon, 2005). If the aim of language teaching is to help rich networks to grow in the learners' minds, the benefits of explicit teaching are very clear. On the one hand, it compensates for the rich input that an L2 learner lacks by guiding the learner to accurate generalizations; and on the other hand, it provides the richly varied range of experiences that a learner needs to embed each new word in a distinct and rich network (3.1).

Incidentally, I notice that the hyperpolyglots discussed in 1.2 learned most of their languages from books, i.e. by learning general patterns explicitly; and indeed these short cuts were presumably essential for their prodigious achievements.

4.2 How to make language interesting

Finally, we come back to the question of motivation (3.3). What can schools and teachers do to motivate learners? Everyone knows that this is the crucial variable in L2 teaching, so it deserves attention from all concerned and I believe even theoretical and descriptive linguistics may have something to contribute.

As I suggested in 3.3, this problem is especially acute when the learners have neither of the two classic types of motivation, integrative and transactional, which is surely the case in the UK and probably in most other English-speaking countries as well. The typical language-learner in school is not moved by talk of far-off goals such as being able to get a job or make friends abroad in ten or twenty years time. No other school subject survives on that basis, so it is hard to see why L2 teaching should be different. To make a child look forward to the next French lesson, the benefits must be much more immediate. One way to ensure this is to offer lots of fun in the lesson, and I certainly applaud anything which achieves this – role play, email exchanges with foreign schools and so on. But this kind of motivation risks being counter-productive if it channels all of the learners' attention onto the activity and away from the language.

A much better solution is to make the language itself interesting, because this guarantees that attention will be on the words and their network connections. If the goal is to enrich network connections, it probably doesn't matter how this is done, and there are many ways of doing it – looking for related words in L2 or etymons in L1, playing games with scrambled letters or words, even solving patterns problems as in the Linguistics Olympiad (http://www.philol.msu.ru/%7Eotipl/new/main/mol/index-en.php). The Department for Education and Skills in England has recently produced an impressive list of suggestions for teaching L2 interestingly in primary schools (Anon, 2005), but there are many other sources of ideas – not least, academic linguists, as people who almost by definition are driven by interest in language. Even more importantly, every academic linguist I know was already fascinated by language as a school child, so we all have personal experience of what is needed. If every L2 teacher combined cognitive linguistics with all the skills and knowledge of a good language teacher, L2 learners would have a really good deal.

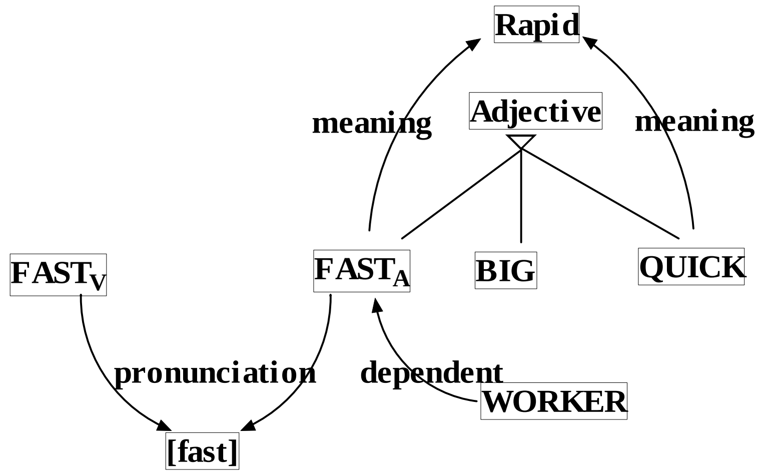

Figure 1: A tiny network centred on the adjective FAST

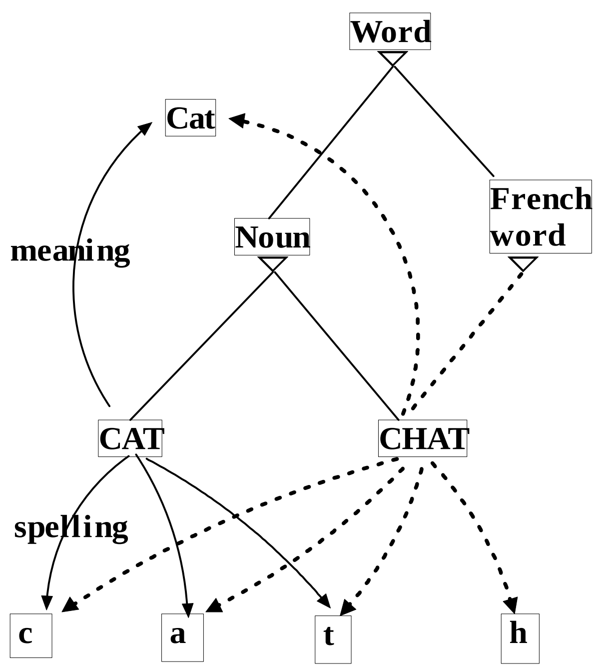

Figure 2: An English speaker learns the French word *chat*

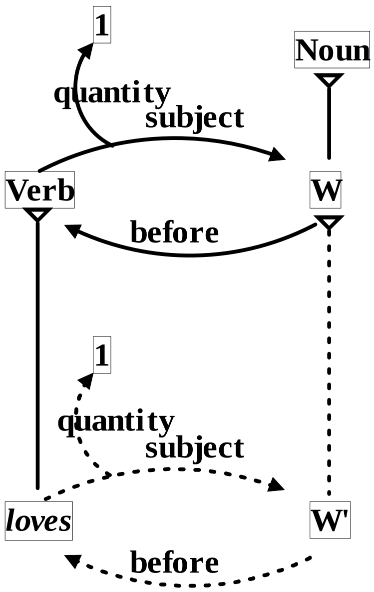

Figure 3: ***Loves*** **inherits its subject properties from Verb**

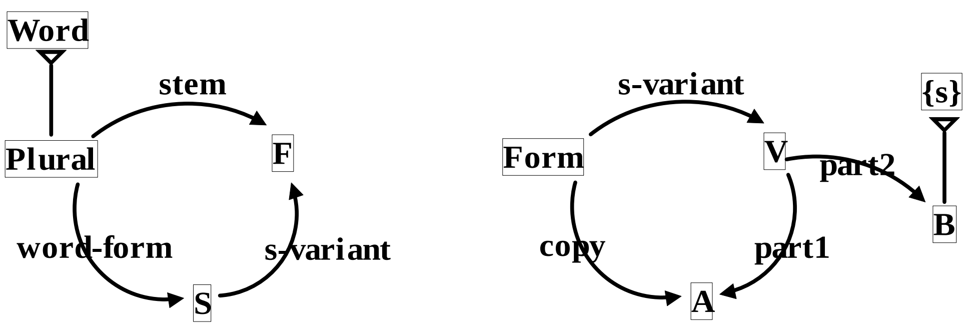

Figure 4: The morphology of plural nouns

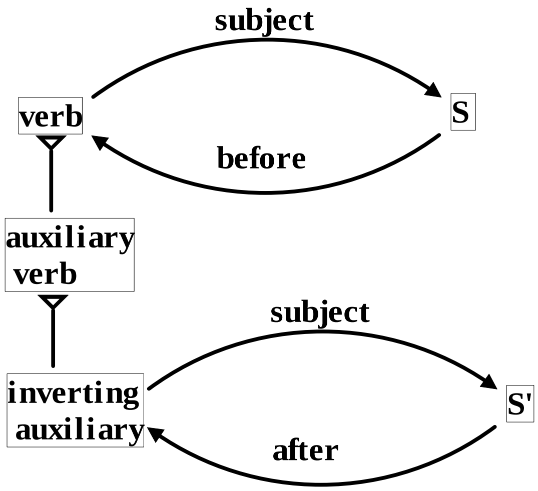

Figure 5: Subject-auxiliary inversion in a network

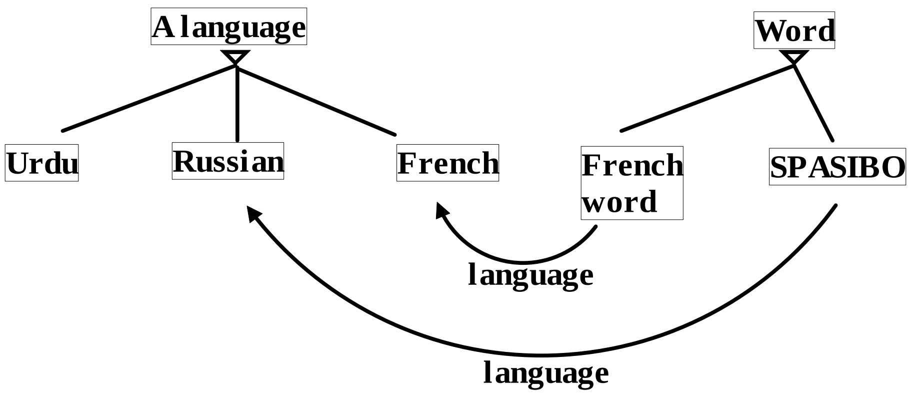

Figure 6: Three languages that I know about and the words I know in them

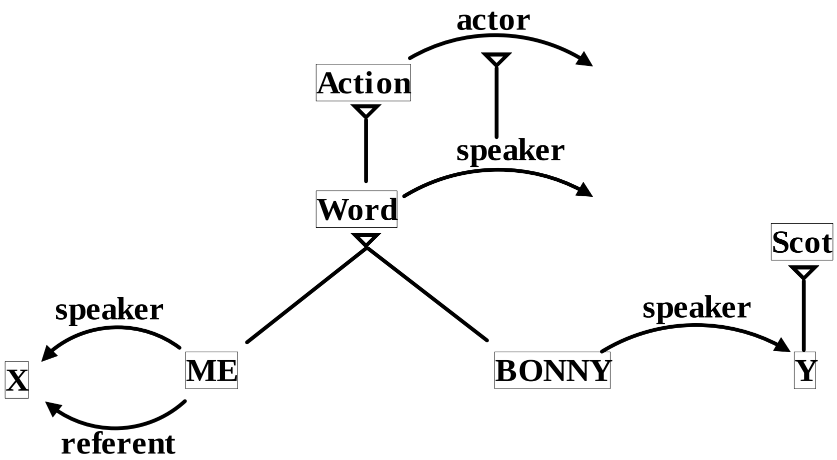

Figure 7: What I know about the speaker of two English words

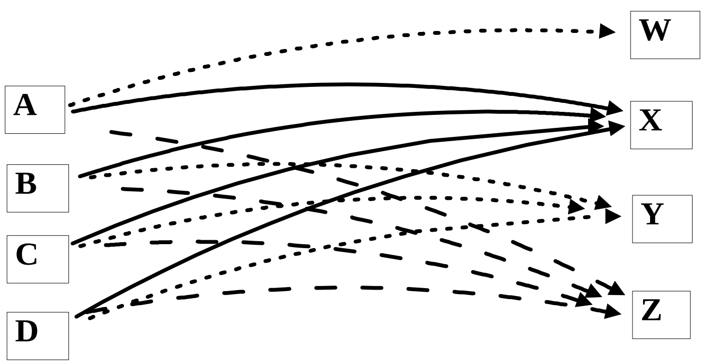

Figure 8: an abstract network showing three correlated properties

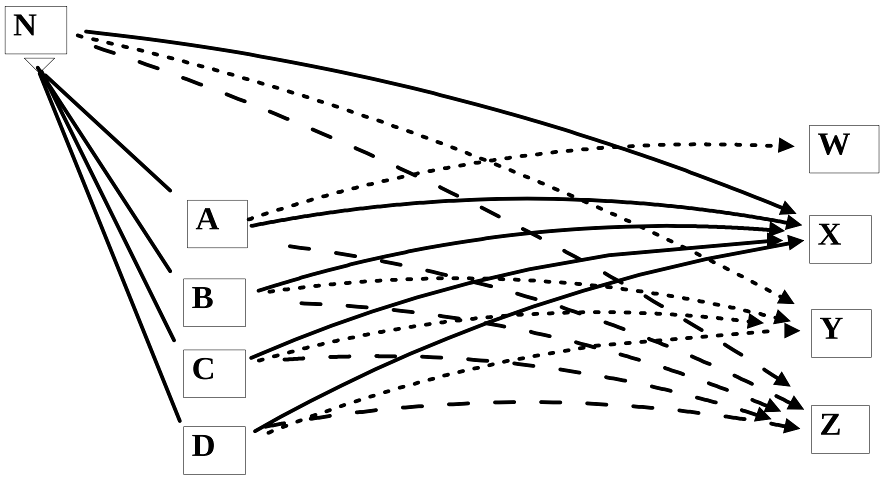

Figure 9: A new node carries the default properties

Reference List

Aitchison, J. (1997). *The language web*. Cambridge: Cambridge University Press.

Aitchison, J. (1994). *Words in the Mind. An Introduction to the Mental Lexicon. Second Edition.* (2 ed.) Oxford: Blackwell.

Anon (1998). *The Grammar Papers*. London: Qualifications and Curriculum Authority.

Anon (2002). *Languages for all: languages for life. A Strategy for England.* London: Department for Education and Skills.

Anon (2005). *Key Stage 2 Framework for Languages*. London: Department for Education and Skills.

Barlow, M. & Kemmer, S. (2000). *Usage Based Models of Language*. Stanford: CSLI.

Bock, K. & Griffin, Z. (2000). The persistence of structural priming: Transient activation or implicit learning? *Journal of Experimental Psychology-General, 129,* 177-192.

Bybee, J. (1999). Usage-based phonology. In M.Darnell, E. Moravcsik, F. Newmeyer, M. Noonan, & K. Wheatley (Eds.), *Functionalism and formalism in Linguistics. I: General papers* (pp. 211-242). Amsterdam: Benjamins.

Bybee, J. & Hopper, P. (2001). *Frequency and the Emergence of Linguistic Structure.* Amsterdam: Benjamins.

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

Crystal, D. (1997). *Cambridge Encyclopedia of Language, 2nd edition*. Cambridge: Cambridge University Press.

Ellis, N. (1994). *Implicit and Explicit Learning of Languages.* . London: Academic Press.

Ellis, N. (2002). Frequency effects in language processing: a review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition, 24,* 143-188.

Erard, M. (2005). The gift of the gab. *New Scientist, 2481 (January 8),* 40-44.

Fodor, J. (1983). *The Modularity of the Mind*. Cambridge, MA: MIT Press.

Gardner, R. & Lambert, W. (1972). *Attitudes and motivation in second language learning*. Rowley, MA: Newbury House.

Giles, H. & Bradac, J. (1994). Speech styles: attitudes and inferences. In R.Asher (Ed.), *Encyclopedia of Language and Linguistics* (pp. 4260-4264). Oxford: Pergamon.

Giles, H. & Powesland, P. (1975). *Speech Style and Social Evaluation*. London: Academic Press.

Goldberg, A. (1995). *Constructions. A Construction Grammar Approach to Argument Structure.* Chicago: University of Chicago Press.

Harley, T. (1995). *The Psychology of Language*. Hove: Psychology Press.

Hawkins, E. (1999). Foreign language study and language awareness. *Language Awareness, 8,* 124-142.

Hudson, R. (1976). *Arguments for a Non-transformational Grammar.* Chicago: Chicago University Press.

Hudson, R. (1984). *Word Grammar.* Oxford: Blackwell.

Hudson, R. (1990). *English Word Grammar.* Oxford: Blackwell.

Hudson, R. (1996). *Sociolinguistics, 2nd edition*. Cambridge: Cambridge University Press.

Hudson, R. (1997). Inherent variability and linguistic theory. *Cognitive Linguistics, 8,* 73-108.

Hudson, R. (2000). Language as a cognitive network. In H.G.Simonsen & R. T. Endresen (Eds.), *A Cognitive Approach to the Verb. Morphological and Constructional Perspectives* (pp. 49-72). Berlin: Mouton de Gruyter.

Hudson, R. (2002a). Richard Hudson. In K.Brown & V. Law (Eds.), *Linguistics in Britain: Personal Histories* (pp. 127-138). Oxford: Blackwell.

Hudson, R. (2002b). Word Grammar. In K.Sugayama (Ed.), *Studies in Word Grammar* (pp. 7-32). Kobe: Research Institute of Foreign Studies, Kobe City University of Foreign Studies.

Hudson, R. (2003). Mismatches in Default Inheritance. In E.Francis & L. Michaelis (Eds.), *Mismatch: Form-Function Incongruity and the Architecture of Grammar* (pp. 269-317). Stanford: CSLI.

Hudson, R. (2004). Why education needs linguistics (and vice versa). *Journal of Linguistics, 40,* 105-130.

Hudson, R. (2007). *Networks of language: the new Word Grammar*. Oxford: Oxford University Press.

Hudson, R. & Walmsley, J. (2005). The English Patient: English grammar and teaching in the twentieth century. *Journal of Linguistics, 41,* 593-622.

Kelly, M. & Jones, D. (2003). *A New Landscape for Languages.* London: Nuffield Foundation.

Labov, W. (1989). The child as linguistic historian. *Language Variation and Change, 1,* 85-97.

Lamb, S. (1966). *Outline of Stratificational Grammar*. Washington, DC: Georgetown University Press.

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

Langacker, R. (2000). A dynamic usage-based model. In M.Barlow & S. Kemmer (Eds.), *Usage-based Models of Language* (pp. 1-63). Stanford: CSLI.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral And Brain Sciences, 22,* 1-45.

Lewis, R. (1996). Interference in short-term memory: The magical number two (or three) in sentence processing. *Journal of Psycholinguistic Research, 25,* 93-115.

Meara, P. (1998). Second language vocabulary acquisition: A rationale for pedagogy. *Applied Linguistics, 19,* 289-292.

Miyake, A. & Shah, P. (1999). *Models of Working Memory. Mechanisms of active maintenance and executive control.* Cambridge: Cambridge University Press.

Muysken, P. (2000). *Bilingual Speech. A Typology of Code-Mixing.* Cambridge: Cambridge University Press.

Norris, J. & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning, 50,* 417-528.

Pinker, S. (1998). Words and rules. *Lingua, 106,* 219-242.

Reisberg, D. (1997). *Cognition. Exploring the Science of the Mind.* New York: Norton.

Renou, J. (2001). An Examination of the Relationship between Metalinguistic Awareness and Secondlanguage Proficiency of Adult Learners of French. *Language Awareness, 10,* 248-267.

Smith, N. V. (1999). *Chomsky. Ideas and ideals*. Cambridge: Cambridge University Press.

Tomasello, M. (2003). *Constructing a language: a usage-based theory of language acquisition.* Harvard University Press.

Wray, D. (1994). *Language and Awareness*. Hodder and Stoughton.