Rolf Kreyer

The Nature of Rules, Regularities and Units in Language

Cognitive Linguistics Research

Volume 51

Rolf Kreyer

The Nature of Rules, Regularities and Units in Language

A Network Model of the Language System and of Language Use

DE GRUYTER
MOUTON

ISBN 978-3-11-031832-6
e-ISBN 978-3-11-031871-5
ISSN 1861-4132

Library of Congress Cataloging-in-Publication Data
A CIP catalog record for this book has been applied for at the Library of Congress

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.dnb.de.

Printing: Hubert & Co. GmbH & Co. KG, Göttingen
♾ Printed on acid-free paper
Printed in Germany

www.degruyter.com

to Anna

Acknowledgements

There are numerous people from whom this book has greatly benefited. I am indebted to them for their help and comments and I am glad to have the opportunity to thank some of them in particular. First and foremost, I would like to thank Jürgen Esser who read earlier versions of the manuscript and provided detailed comments and most invaluable advice. Without him this book would never have been completed.

I would also like to thank Hannes Kniffka, Manfred Kohrt, Karl Reichl, Klaus Peter Schneider and three anonymous reviewers who read an earlier draft and provided me with very helpful and detailed feedback. For insightful comments and stimulating discussions on individual aspects I am grateful to Barbara Güldenring and Steffen Schaub. To the former special thanks are also due for proof-reading the whole manuscript. I would also like to express my gratitude to the series editor, Dirk Geeraerts, for including this book in *Cognitive Linguistics Research*. Naturally, I am responsible for all remaining errors and infelicities.

Finally, I would like to express my deepest gratitude to my beloved wife, Anna, for her unfailing and continuous support and love. To her this book is dedicated.

Content

1 Introduction

Richard Hudson, in his many publications within the field of Word Grammar, formulates what he refers to as:

> **The Network Postulate**
> Language is a conceptual network (Hudson 2007a: 1, among others).

This view, as Hudson claims, can be regarded as a "commonplace of modern linguistics" (1) if we take into consideration the structuralist idea of language as a system in which each element obtains its value from its relation to other elements in the system: "Any system of interconnected entities is a network under the normal everyday meaning of this word" (1).

Uncontroversial as this claim may appear at first sight, problems may arise if the claim is taken seriously, i.e. "language is **nothing but** a network – there are no rules, principles, or parameters to complement the network. Everything in language can be described formally in terms of nodes and their relations" (1; Hudson's emphasis). Such a view of language, although in line with cognitive approaches and psycholinguistic models of language, runs counter to basic tenets of traditional linguistic schools, such as a strict division between lexis and grammar or, more generally, between the fully regular and the idiosyncratic in language. Network models do away with this neat bipartition: "The claim that language is a network therefore conflicts with the claim that information is divided between the grammar and the lexicon. In a network analysis, the same network includes the most general facts ('the grammar') and the least general ('the lexicon'), but there is no division between the two" (Hudson 2007a: 3). This example illustrates a very important point: "[T]he conceptual-network idea is not merely a matter of our choice of metaphors for thinking about language or what kinds of diagram we draw. It also has important consequences for the theory of language structure […]". (Hudson 2007a: 4)

The model of language suggested in the present study is in line with the general claims underlying the models suggested by Hudson and others. It follows the Network Postulate, in that it understands language to be a network and nothing but a network. However, the present study wants to take the network idea one step further. The appeal of network models lies in the fact that the brain itself consists of a network of nerve cells and links between these nerve cells. Network models seem to be inherently closer to a possible psychological or neurophysiological reality and, hence, more cognitively plausible. However, existing network models of language (apart, perhaps, from purely connectionist models of the kind illustrated in MacWhinney et al. 1989) make use of a descriptive apparatus which goes beyond what brain cells and neurophysiology have to offer, e.g. different kinds of links between nodes (such as ISA or relation nodes, e.g. 'giver' or 'receiver'; Hudson 2007a among others), numbered links to represent sequence of elements (Roelofs 1997) or different classes of nodes to represent concepts like AND or OR (Lamb 1998). The present study, therefore, is an attempt to meet what could be referred to as:

> **The Neurophysiological Challenge**
> Language is nothing but a network. This network mirrors the neurophysiology of the brain in that the nodes and the connections are closely modelled on what we know from nerve cells and their connections.

The 'ingredients' of the network are, thus, significantly reduced: the network consists of nodes, which represent sounds, morphemes, syntactic structures, etc. These nodes can be activated to different extents. If this activation meets a certain threshold level, the node will pass on activation to all the nodes to which it is connected. A higher degree of activation may lead to a faster passing on of activation. Activation is passed on through links between nodes. These links do not represent relations; rather they are of an associative kind. Links just serve as conduits through which activation can be transferred. They always start in a node but they may end in a node as well as in another link. In line with basic neurophysiology, links are of two kinds: 1) excitatory links increase the activation of the target, while 2) inhibitory links decrease or block the activation of the target (see also Lamb 1998).

That is the basic machinery underlying the present network. In addition to these ingredients, network structures may change in different ways, thus implementing the view advocated in Bybee (2010: 2; among others): "language can be seen as a complex adaptive system [...], as being more like sand dunes than like a planned structure, such as a building". The frequent activation of a node will lead to a lowering of the activation threshold that this node has. That is, frequently 'used' nodes will be accessed and activated more easily in future. Similarly, links that are used more frequently will become stronger, so that they pass activation on more quickly. In general, portions of the network that are used a lot become trained so that future use is facilitated. In this way, the network implements the cognitive notion of entrenchment (e.g. Langacker 2008). In addition, nodes grow links between one another if they are co-activated frequently. This is the network correlate of association of previously unrelated

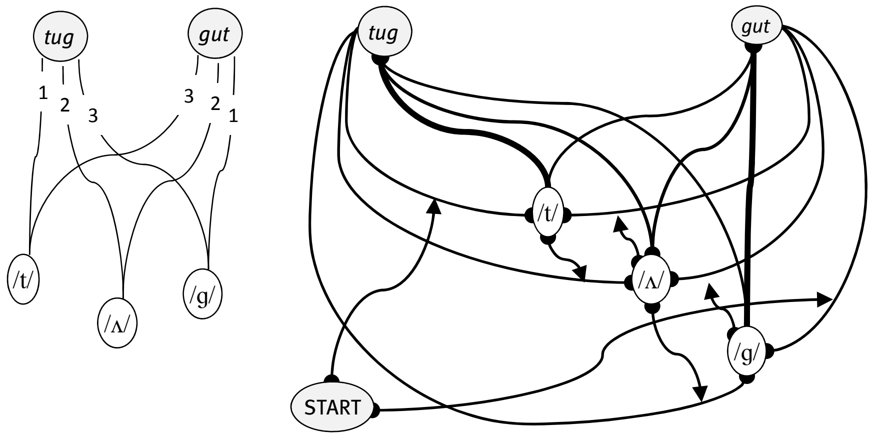

Figure 1.1: Sequences in a network model with ranked (left) and purely associative links (right).

events; it explains well-documented phenomena like priming and important concepts like collocation or colligation. Finally, the network may 'grow' new nodes as the result of new experiences or the result of categorization.

In following these neurophysiological restrictions the present model is in line with the idea of 'cognitive commitment', i.e. "a commitment to make one's account of human language accord with what is generally known about the mind and brain, from other disciplines" Lakoff (1990: 40). This commitment has far-reaching consequences for the network model. For instance, simple linguistic facts, like the order in which, say, phonemes appear in a word, are extremely easy to represent in a network model that allows for links that express ranks. To achieve the same is a lot more difficult in a network that is based on neurophysiological facts. Instead of ranked links, such a model has to apply links of different strengths, the notion of word beginnings and an intricate pattern of excitatory connections (see figure 1.1).

The obvious question, then, is: 'What is the advantage of a significantly more complicated way of representation in the present model?' The first answer is that this way of representation is more accurate in that it follows what we know from neurophysiology. If in the brain we do not have axons (the connection leading from a neuron to another neuron) that can represent ranks, a network model should try to make do without such a kind of connection. The second answer is related to that. The model presented in this study, similar to the suggestions made by Hudson or Lamb (among others), is of the hard-wired type. It is not self-organising but it is the researcher that decides on the number and kind of nodes and on the connections that exist between these nodes. In connectionist circles, this approach has been dismissed as being of little value, since it is assumed that more or less everything can be hard-wired. This would, then, make models like those presented by Hudson or Lamb and the present model little more than mere presentational variants of other, more traditional kinds of representation. There is no need, here, to go into further detail regarding this aspect. Let it suffice to say that, as we have seen above, existing network models, by their very nature, do make claims about language structure that go far beyond or even contradict traditional models. Again, the present study would like to take this development one step further. It is not only concerned with how particular facts of language are represented in a network but it also tries to account for the processes that are going on in the network as soon as particular nodes are activated. Following Langacker (1987: 382), language in the present network model "is not something a speaker *has*, but rather what he *does*." This is an additional challenge and leads to problems that are far from trivial. For instance, it is fairly easy to describe an activation pattern that represents the fact that in the sentence *John reads books* the finite verb inherits or copies person and number features from the subject *John* and not from the patient *books*. But it is fairly difficult to design a network structure which leads to the correct activation pattern as soon as the network 'knows' that *John* is the agent of the action of reading. In particular, how can we make sure that the plurality of the patient does not interfere with this process? Without going into further detail at this point (see section 4.2.4 for a detailed discussion), the structure that solves this problem is shown in figure 1.2.

The present study, thus, not only presents an alternative to network models like those of Hudson or Lamb, it also tries to provide a model that explains how network structures interact in the production and the comprehension of language. The model is a model of the static language system but, at the same time, accounts for the system in action. In this, the present study follows Lamb's (2000: 95) dictum that a model of the language system "must represent a competence to perform". Which nodes to include and which connections to establish depends on the question whether such changes increase the performance of the whole system or not. Finally, the present study seeks to explain how network structures evolve on the basis of two fundamental cognitive processes, namely association and categorization. Association is a very basic process that underlies many cognitive operations. As mentioned above, it refers to the fact that events (or rather: stimuli) that co-occur frequently come to be seen as related and interdependent in the cognitive system of the organism that experiences these events. Association lies at the heart of categorization, since categories are

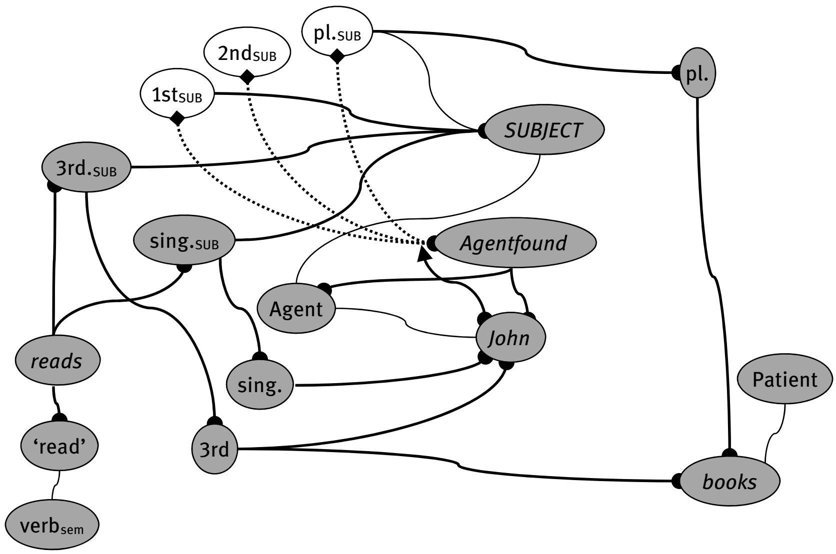

Figure 1.2: Choosing the right verb form if the agent *John* is the subject of the clause *John reads books*. (details omitted).

understood as an accumulation of features that co-occur frequently. A semantic category, for instance, is described as a set of co-occurring semantic features. This view of categories makes it very easy to implement what Langacker (2000b: 4) refers to as 'schematization', the "capacity to operate at varying levels of 'granularity' (or 'resolution')." Different degrees of abstractness or specificity can be represented and accessed simply by manipulating the number of features that are seen as relevant for membership in a given category. This kind of flexibility is indispensable given the enormous diversity of categories in language. Categories are highly useful in interpreting and organising new data, since they can help to make apparent patterns that otherwise would remain hidden: the two sentences *I like books* and *The boy kissed the girl* are as distinct from each other as they can be. It is only through viewing them with regard to functional categories that it is revealed that both sentences are instantiations of the same clause pattern, namely SVO. From this example it becomes apparent that the language system unfolds through an iterative process of association, which leads to categorization, which leads to new kinds of association, which leads to new categories and so on.

This process is fundamental to the description of any language system. The aim of the present study is not merely to show that language systems (in this case the English language system) can be described by the present network model. One of its other major aims is to show that the iterative process described above is a consequence that follows from the nature of the network and the processes going on in the network. The network model, therefore, has a high degree of explanatory power: the structures that we find in the network are the result of the nature of the network itself, i.e. they are the result of elements and processes that are similar to the elements and processes that are found in the human brain. Similarly, it follows from the nature of the network that basic-level categories are optimal for processing. As a consequence, from a network point of view, abstract notions like 'XP', although they may increase the elegance of academic descriptions of language, do not seem to be relevant. Other aspects that can be derived from the nature of the network include redundancy of storage, i.e. the fact that complex forms are stored if they are frequent enough or the fact that the language system contains structures at different levels of abstraction and specificity that transcend rank boundaries. Similarly, the nature of the network itself suggests that the model is both local and distributed at the same time. That is, some concepts are represented by a single node in the network, while others are represented by the co-activation of a number of different nodes. In this way, the present model does not fall prey to the exclusionary fallacy that networks either have to be distributed or local, or, for that matter, that they have to be real-copying or virtual copying – the nature of the network suggests that both kinds of information-inheritance co-exist. Finally, in some cases the network suggests answers to problems of linguistic description. An example is the problem of gerunds in English. Without going into detail at this point (see section 4.2.2), the present network would rule out a solution where an –*ing* form is classified as a verb and a noun at the same time. In cases like these, the study (or its author) does not choose a particular representation of a linguistic feature but the study shows what solutions are possible if we consider the assumptions underlying the network as valid.

On the whole, the present study suggests a network model of language that restricts itself to elements and processes known from the human brain. The model should not be understood as a presentational variant of other models of language – network or other. Rather, the study wants to show how a system that is based on neurophysiological 'ingredients' can create structures that represent units and structures of language; at the same time, the study wants to show how these structures, when in operation, lead to activation patterns that represent the outcome of processes of language production and comprehension as described over the last few decades. This way, the present study hopes to make a contribution to the cognitive linguistic enterprise.

The study is organized in six chapters. After the introductory remarks in this chapter, chapter 2 discusses a number of standards that have to be met for a model of language to be regarded as cognitively plausible and it introduces the basic features of the network model advocated in this study. Chapter 3 discusses the steps of any linguistic theory from data to description and from description to the formulation of rules and shows how these steps are represented in the evolution of network structures. The following three chapters explore the implementation of a wide range of linguistic phenomena in the network model. Chapter 4 focuses on the description of traditional concepts and notions in the present model including the representation of sequences, structures and rules. It also discusses aspects like gradience, redundancy and ambiguity. Chapter 5 looks at more recent concepts discussed in cognitive and corpus-linguistic circles, namely cognitive schemas (including aspects of construction grammar) and various kinds of multi-word units. Chapter 6 focuses on language use. It explores how the nature of categories and aspects like expectancies are exploited during processing and how the network structures relate to claims made in research on processing principles such as the end-weight principle.

2 A cognitively plausible network model of the language system

This chapter introduces, in the first part, a number of requirements that have to be fulfilled for any language model to be cognitively plausible and that the present model is aiming for. In part, these requirements are based on considerations of a more general kind. Some requirements, however, are formulated envisaging the vast array of linguistic concepts and phenomena that a comprehensive model has to encompass. In this sense, the first part of this chapter can also be understood as a preview of chapters 3 to 6. The second part develops a network model of the language system that meets all of these demands.

2.1 A cognitively plausible model

2.1.1 A usage-based model

'Usage-based' can be understood in two ways, namely as relating to quantities that are found in language use and their representation in a model of language or as relating to an empiricist view. Although throughout this study there will be references to frequencies and statistical associations, 'usage-based' is primarily interpreted in an empiricist way: all the structures and elements that are described in this model are understood as being based on language data that the language user has encountered. Every structure that will be described in the remaining chapters is thought of as resulting from the application of general cognitive mechanisms to these language data. More specifically, the model is informed by Langacker's (1988a) 'content requirement' and thus ensures that all the elements in the description can be traced back to word forms or strings of word forms that actually occur in language data. That is, the model only contains elements that are of one of the following kinds: 1) elements that are direct representations of overtly occurring language strings, i.e. representations of word forms and of strings of word forms. 2) abstractions or schemas that results from generalizations of these (strings of) word forms. A category like 'past tense', for example, is the result of schematization on the basis of a set of actually occurring past tense forms like *kissed*, *begged*, *loved*, *hugged* and so on. 3) relations that exist between overtly occurring language strings, between abstractions and overtly occurring strings, or between abstractions and abstractions. An example of the first kind of relation is that of collocation, e.g. the co-occurrence of the two word forms *naked* and *eye*. Relations between abstractions and overtly occurring strings are exemplified by the class 'regular past tense verbs' and all of its instantiations. Finally, an example of relations between abstractions and abstractions is the instantiation relation between clause types such as the mono-transitive clause and its instantiating patterns SVO and OSV. It is clear that the general usage-basedness of the present model rules out the concepts developed in more recent approaches in the generative-transformational paradigm, since "[t]he notion of grammatical construction is eliminated" (Chomsky 1995: 170) in these theories (see section 4.1).

2.1.2 A redundant-storage model

Given the usage-basedness it makes sense to assume that the model stores information redundantly. This is in line with many cognitive accounts of language (and also other network models, e.g. Hudson 2007a) and is suggested by findings from different areas of language research. One of these is language acquisition. For instance, Tomasello and his co-workers find that child language acquisition starts off with "specific item-based constructions" (Tomasello 2006: 285) which only later become more abstract. It is important to note, though, that "as children create more general constructions, they do not throw away their more item-based and local constructions" (Tomasello 2003b: 11). In this way, rules or abstract constructions and concrete instantiations are supposed to co-exist, at least in the mental grammar of the child during language acquisition.

In a similar vein, Langacker (1987: 28) argues against what he calls the 'exclusionary fallacy', an assumption which wrongly suggests "that one analysis, motivation, categorization, cause, function, or explanation for a linguistic phenomenon necessarily precludes another." A special instance of the exclusionary fallacy is the 'rule/list fallacy', i.e.:

> the assumption, on grounds of simplicity, that particular statements (i.e. lists) must be excised from the grammar of a language if general statements (i.e. rules) that subsume them

> can be established. Given the general N + -s noun-pluralizing rule of English, for instance, specific plural forms following that rule (*beads, shoes, toes, walls*) would not be listed in an optimal grammar. (Langacker 1987: 29)

Recent research rejects such a view on the basis of its lack of psychological plausibility, as Dabrowska (2004: 27) points out: "[h]uman brains are relatively slow processors, but have enormous storage capacity. From a psychological point of view, therefore, retrieval from memory is the preferred strategy". Langacker (2000b) integrates storage as well as rules in the following "viable alternative" to the 'rule vs. list' approach, namely:

> to include in the grammar both rules and instantiating expressions. This option allows any valid generalizations to be captured (by means of rules), and while the descriptions it affords may not be maximally economical, they have to be preferred on grounds of psychological accuracy to the extent that specific expressions do in fact become established as well-rehearsed units. Such units are cognitive entities in their own right whose existence is not reducible to that of the general patterns they instantiate. (Langacker 2000b: 2)

Another reason for such a redundant representation arises from findings in the field of corpus linguistics. The study of vast amounts of language-use data shows that many collocations and colligations do not depend on lexemes but on individual word forms, as can, for instance, be seen in Sinclair's (1996: 84) statement that "*blue* and *brown* collocate only with *eyes*", not with *eye* or any of the other forms of the lexical unit (see section 5.3). It therefore makes sense that the word form *eyes* needs to have its own representation in the network (not just as a product of the general rule of plural formation) to do justice to that fact.

On the other hand, not all word forms of a lexeme are equally frequent with respect to their potentiality of use. A case in point is Esser's (2000a) observation that one meaning of TREE, namely 'drawing', has a strong tendency to occur with the singular form only, while such restrictions are not observable with the 'plant' meaning of TREE. Such facts are best represented if word forms are integrated in the network even though they are completely regular and thus may be generated by a grammatical rule. This redundancy is not restricted to the lexical level but may be found at any level of description.

Finally, redundant storage seems to be licensed by experimental data that suggest a strong influence of frequency in language use and the shape of the language system (see section 4.2.6). A case in point is the observation that highly frequent instantiations of regular phenomena are stored even though they could be generated on the basis of the underlying rule they instantiate (see, among others, Stemberger and MacWhinney 1988; Baayen et al. 1997; Gordon and Alegre 1999). As Bybee and Thompson (1997) remark: "[t]he effects of frequency have important implications for our notions of mental representation. There is not necessarily just one representation per construction; rather, a specific instance of a construction, with specific lexical items in it, can have its own representation in memory if it is of high frequency" (Bybee and Thompson 1997: 386).

On the basis of this overview of facts and previous findings it seems reasonable to demand of any theory of language that it is able to implement redundancy on any level of the language system.

2.1.3 A frequency-based model

Frequency, in addition to redundant storage, figures prominently in other areas of linguistic description. Over the last one or two decades the importance of frequency in the shaping of the language system has been underlined by many researchers (see section 4.2.6 for a more detailed account). According to some, frequency effects are all-pervasive and can be found in any area of language processing and production. Ellis (2002), for instance, claims that there are "frequency effects in the processing of phonology, phonotactics, reading, spelling, lexis, morphosyntax, formulaic language, language comprehension, grammaticality, sentence production, and syntax" (Ellis 2002: 143).[1]

Corpus-linguistic and cognitive linguistic research has emphasised the role that frequency and statistical associations play in the shaping of the language system. Firstly, we witness many, usually frequency-based, patterns of co-occurrence that are characteristic of language use, such as collocations, idioms, or lexical phrases. Secondly, the combinatorial potential of the language system is not exploited to its full extent in language use: For instance, the different meanings of polysemous items often differ significantly in the frequency of realization, frequencies of words and constructions change in dependence of the genre. These differences in frequency influence the processing of language and the shape of the language system itself (see section 6.2). Highly frequent units and strings are more deeply entrenched in the language system[2]: highly frequent items are retrieved faster and more accurately than less frequent items, and highly frequent strings are reported to coalesce over time and to be stored

1 See, for instance, the edited volume by Bybee and Hopper (2001), which documents frequency effects on many levels of linguistic description.

2 Of course, frequency is not the only source of entrenchment. Other possible sources include (partial) opacity of meaning or a particular pragmatic function (see section 5.5).

as one unit even if they are fully regular and thus analysable into their component parts. A plausible model of the language system will need to take these aspects into account. The present model meets this requirement, albeit in a fairly rudimentary fashion by implementing different degrees of entrenchment of nodes and connections. These are designed to capture crude differences but are not meant to mirror correlations or conditional probabilities precisely.

2.1.4 A comprehensive model

Previous attempts at modelling the language system have often been restricted by focusing on some levels of linguistic description only, while at the same time ignoring others. This can be witnessed in the large number of dichotomies and distinctions usually drawn in traditional linguistics, such as those between lexis and grammar, between decontextualized and contextualized units, or between the semantics of sentences and the pragmatics of utterances. Drawing these distinctions, no doubt, has been useful and has led to many important insights. Yet, although a number of aspects concerning the language system can be studied in isolation from other areas of language description, this should not lead us to ignore the fact that many aspects of the language system show interdependencies transcending borders that are usually drawn. This becomes most obvious with regard to the distinction of lexis and grammar, which, according to more recent corpus-linguistic and cognitive research, can no longer be upheld (see the discussion of patterns and pattern grammar in section 5.1[3]).

Other points become apparent as soon as we take seriously the assumption that "[t]he mental system [...] is not some kind of abstract 'competence' divorced from performance, but a competence to perform" (Lamb 2000: 94). That is, we may assume that the language system is based on and shaped by language use. It, therefore, seems reasonable to additionally include in a description of the language system those aspects that are usually relegated to the study of language use and that are mostly regarded as being (at best) only indirectly relevant for the study of the language system. One consequence that arises from the consideration of competence as "a competence to perform" concerns efficiency of language processing and production. It is safe to assume that a model of the language system should be geared towards efficiency (see also the discussion in 6.3), and an efficiency-driven processor will make use of any bit of information available to ensure smooth and accurate interpretation of language data. For

3 In this respect also see Hasan (1987) and her notion of lexis as most delicate grammar.

instance, we are all aware of the effect of genre on the interpretation of utterances and on the disambiguation of word senses, since some word senses are more frequently used in particular genres or discourse types. It stands to reason that a processor will make use of such information. Hence, a plausible model of the language system should integrate contextual information where possible.

Since the envisaged model tries to be comprehensive with regard to these, usually isolated, areas of linguistic description, it will show a wide range of 'elements' that are included in the descriptive apparatus: The model is based on actually occurring strings of word forms, but it will also include information that can be abstracted from these strings of word forms (see the discussion of 'a usage-based model' above). Such pieces of information, for instance, may be a particular semantic feature, it may by a syntactic function like subject or direct object or structures like that of phrases or clauses. The model will also incorporate elements that make reference to the situation of use, such as spontaneous speech or edited writing, and the genre, such as written academic or prose fiction. In short, the envisaged model is comprehensive in that it takes into consideration all aspects of language system and language use simultaneously.

2.1.5 An integrative model

Related to the above is the problem that linguistic description, for decades of research, has upheld the distinction between what is regular and irregular, between productive rules and repeated idiomatic formulae, or between core and periphery. Such distinctions are usually not easy to draw: the division between the regular and the irregular is often a highly subjective one – when is the proportion of regular items large enough to justify the stating of a rule? Also, research by Eleanor Rosch and her colleagues (e.g. 1975) points out that the categories of linguistics are not Aristotelian but rather of a prototypical kind that exhibit what has been referred to as fuzziness, gradience or indeterminacy. The distinction of the productive versus the idiomatic is similarly problematic. The number of completely fixed idioms is rather small, while the largest share of idiomatic expressions show productive variation (see Barlow 2000 for examples). On the other hand, recent corpus linguistic research has made clear that what traditionally has been assumed to be the result of productive application of rules, to a large extent, is based on the use of prefabs or formulae and other kinds of more idiomatic strings or units. Erman and Warren (2000), for instance, claim that more than half of a given text consists of prefabricated phrases. On the whole, there is no language-inherent way of drawing a clear distinction between a core characterized by full generality, productivity and regularity and a periphery of completely particular, idiomatic and irregular items and strings. Rather, we are dealing with a continuum of different degrees of 'coreness' or 'peripherality', which suggests one descriptive apparatus to account for all phenomena found in the study of language (see also the discussion of cognitive schemas in section 5.2).

2.1.6 A hierarchical model

No model of language can dispense with different levels of hierarchy in its descriptive apparatus. One reason for this is the phenomenon of 'constituency', i.e. the observation that a particular linguistic item uses items on a lower level at its building blocks. In many cases, linguistic rules are nothing but a formulation and description of different relations of constituency in language data. In addition, hierarchies are demanded by the fact that we can make use of different degrees of detail in our description of linguistic phenomena.[4] Different degrees of detail automatically lead to a hierarchical structure where a superordinate and less differentiated class encompasses a number of subordinate and more differentiated classes. A similar intuition also becomes apparent in Rosch et al.'s (1976) distinction of 'superordinate', 'basic-level', and 'subordinate' categories (see section 6.1). The advantage of such a system lies in the fact that it enhances processing by enabling the language system to work with those amounts of information that are needed at a particular point in time and not burden itself with additional information that would be useless at that particular moment (see also the discussion of schemas in section 5.1). This feature can be witnessed in many different areas of cognition. Consider, for instance, people at a ball. If a new song begins, dancers might at first only be interested in the rhythm in order to determine whether they should dance a waltz, a rumba or a tango. In this situation more detailed information on the song being played (such as the instruments being used, the language in which the singers sing, etc.) might actually impede the recognition of the style of music. When crossing a street, processing on a very low level of delicacy also seems to have its advantages. The most important information is already carried by fairly large and unspecific categories such as 'pedestrian', 'bicycle', 'motorbike', and 'car', since these tell us something about the speed and the potential to harm us when

4 In this respect, see Halliday's (1961: 272) notion of 'delicacy', i.e. "the scale of differentiation, or depth in detail".

crossing the street. More detailed information (brand of car or motorbike, colour, etc.) might make it difficult for us to process the relevant pieces of information as fast as necessary. If we want to buy a car, in contrast, a far more detailed level of scrutiny is appropriate. The cognitive system thus seems to be fitted with the ability to 'zoom in and out' of objects, dependent on the amount of information that is needed in a particular situation and for a particular purpose; Langacker's (2000b) idea of 'schematization' (see section 5.1).

A similar ability is also useful in the processing of language structures. Often the processor does not need all the information contained in a linguistic element to draw relevant conclusions and make useful predictions. A case in point is the occurrence of the definite article in a string of words. On the basis of the definite article alone the processor can tell that the structure about to being processed is an NP (see Kimball's 1973 parsing principle 'New Nodes'). This bit of information will lead to expectations regarding certain other features that are relevant for processing. At the beginning of the clause, this NP is very likely to function as the subject; at least in written English. In spontaneous conversation the mere fact that the processor has encountered the beginning of a full NP is a fairly reliable cue for the fact that the NP is not the subject of the clause, since these are mostly realized by pronouns. In both cases, the processor will be fairly safe to assume that the whole NP makes reference to generic entities or entities already given in the previous discourse, since this is what the definite article signals. This makes clear that the very general category 'definite NP' already enables the processor to draw highly relevant conclusions about the general nature of the linguistic string being processed. In the light of principles like Hawkins' (2004) 'Maximize On-line Processing' (see section 6.3), which suggest that the human parser will ascribe ultimate properties of the language string as early as possible, it makes sense to assume that the language system, similar to other cognitive systems, makes use of different degrees of delicacy or granularity during processing. These degrees can be expressed in a hierarchic system.

2.1.7 A rank-permeability model

Having emphasised the importance of hierarchies in any model of language description, it is also important to stress the interaction between elements on different levels of hierarchy. Obviously, any model of language needs to have a certain degree of permeability between adjacent ranks; otherwise, it could not do justice to the phenomenon of constituency that we find in any traditional model of grammar. However, and more importantly, we also witness interdependencies between levels of grammatical description beyond constituency relations (see section 2.1.1). This becomes clear if we take into consideration cognitive schemas or recent corpus-linguistic concepts (sections 5.2 and 5.3, respectively). Cognitive schemas, as exemplified in Fillmore et al.'s (1988) notion of construction or Hunston and Francis' (2000) concept of pattern show a high degree of inter-rank dependencies. Similarly, the notions of semantic preference and semantic prosody (e.g. Sinclair 1991) combine word forms with semantic features, e.g. the string *naked eye* co-occurs frequently with word forms that make reference to visibility and word forms that contain a semantic aspect of difficulty. In summary, a cognitively plausible model of the language system should allow for associations between any kind of element on all conceivable ranks of linguistic description.

The next section will show how a network model of language is able to meet all of the seven standards discussed above.

2.2 A network model

2.2.1 Network models in psychology and linguistics

The idea of using networks to model cognitive and linguistic processes is not new. In particular, network models have been employed in psychology to account for experimental findings regarding the understanding of sentences or the effect of priming on word form recognition. For instance, Collins and Quillian (1969) suggest a hierarchical network model of semantic memory. Their network encodes properties of objects and classes, and the superset-subset relations between them, i.e. the network is primarily organised on the basis of ISA-relations (a 'subset' is a (kind of) 'superset'), e.g. A canary ISA bird and a bird ISA animal. In addition to the ISA-links, the model also has feature links. A canary, for instance, shows the property 'can sing' and 'is yellow'. Superordinate classes also show distinctive features, such as 'can fly' for *BIRD* or 'has skin' for *ANIMAL*. The model largely is what has been called a 'virtual copying' model (see Goldberg 1995: 74), i.e. the information for subordinate items is only stored in the superordinate nodes[5]. The property 'can fly' is, thus, directly con-

5 Note that this is only the simplest case. Collins and Quillian (1969: 242) make clear "that people surely store certain properties at more than one level in the hierarchy". See also Collins and Loftus (1975: 409) for a discussion of this point.

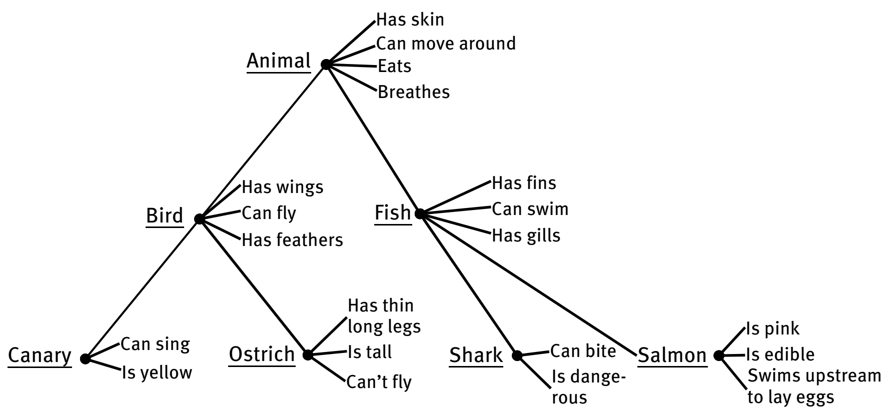

Figure 2.1: The network model of Collins and Quillian (1969: 241) (with kind permission of Elsevier).

nected only to the class *BIRD* but not to the class *CANARY*, and the property 'has skin' is connected to *ANIMAL* but neither to *BIRD* nor *CANARY*. In Collins and Quillian's approach the respective network appears as shown in figure 2.1.

According to this model, it should be easier to verify a sentence like 'a canary can sing' than a sentence like 'a canary has skin', since in the first case the relevant information is directly attached to the 'canary' node, whereas in the second case the relevant information is two nodes away from the 'canary' node. This is exactly what is borne out by Collins and Quillian's data, i.e. sentences of the first kind were verified faster than those of the second kind.

A less strictly hierarchical network model is the one suggested by Collins and Loftus (1975). Their model is based on similarity: "The more properties two concepts have in common, the more links there are between the two nodes via these properties and the more closely related are the concepts" (411). Relatedness, in this model, is expressed by proximity of nodes, as figure 2.2 shows. The different kinds of vehicles in the upper portion of the network are closely related, since they share a number of features. In contrast, 'fire engine' and 'cherries' are not closely related, since they only share the singular feature 'red' and no other features.

Most of the network accounts have focused on restricted aspects of language, such as the mental lexicon (see above and, among others, Beckwith et al. 1991; Fellbaum 1998; Miller and Fellbaum 1991, 1992; Steyvers and Tenenbaum 2005), morphological processes like past tense formation (Rumelhart and

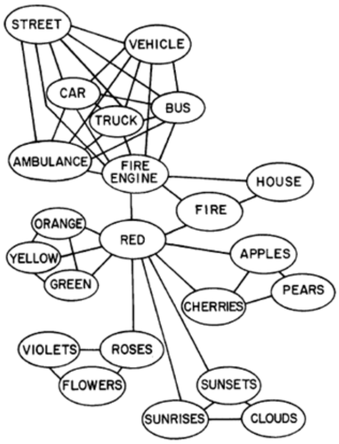

Figure 2.2: The network model of Collins and Loftus (1975: 412) (with kind permission of the American Psychological Association).

McClelland 1986; MacWhinney and Leinbach 1991), or the acquisition of syntactic categories (Gasser and Smith 1998). The only comprehensive network models of the English language, to my knowledge, are provided by Sidney Lamb (1998)[6] and Richard Hudson (2007a, 2007b and 2008). In contrast to the approaches by Collins and Quillian or Collins and Loftus as sketched out above, Lamb makes use of a fairly elaborate system of different kinds of nodes which "differ from one another according to three dimensions of contrast: (1) *UPWARD* vs. *DOWNWARD* orientation, (2) *AND* vs. *OR*, (3) *ORDERED* vs. *UNORDERED*" (Lamb 1998: 66). According to Lamb, the whole language system consists of a network of such nodes related to one another (see figure 2.3).

An upward unordered AND node is shown in the triangle that links 'GO' and 'Verb' to the line next to '*go*'. This means that the activation of this line will activate both 'GO' and 'Verb' and, conversely, that '*go*' will be activated if both 'GO' and 'Verb' are activated. The symbols in the bottom line of figure 2.3 show upward unordered OR links. This means that the phoneme[7] /g/ spreads its activation to all the nodes to which it is connected, i.e. all the forms that contain this phoneme. Conversely, all forms that contain this phoneme will activate the

6 See also Lamb's (1966) first outline of 'stratificational grammar' and Sampson's (1970), Lockwood's (1972) and Schreyer's (1977) introductions to the theory, as well as Makkai and Lockwood (eds.) (1973) for an early collection of papers following Lamb's approach.

7 Despite Lamb's highly idiosyncratic terminology, I will here stick to the traditional terms.

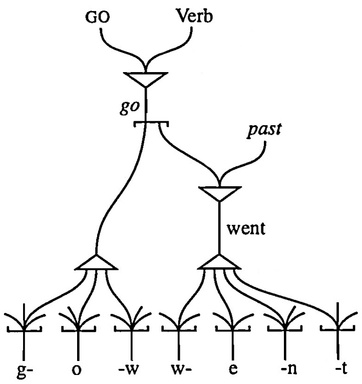

Figure 2.3: A network portion for GO in Lamb's (1998: 60) network (with kind permission of John Benjamins Publishing Company, Amsterdam/Philadelphia).

node representing /g/. As a final example consider the triangle that connects the three phonemes /g/, /o/ and /w/. This triangle denotes a downward ordered AND node: only if the three phonemes are activated in the order in which they are represented in the figure will the node *go* be activated, and, again, the activation of the latter will activate the phoneme nodes in this particular order

Another important point of divergence from the two approaches discussed before resides in the fact that Lamb explicitly denies the necessity of any symbols in the network:

> If the relationships of linguistic units are fully analyzed, these 'units' turn out not to be objects at all, but just points of interconnection of relationships. We may conclude that the linguistic system (unlike its external manifestations) is not in itself a symbol system after all, but a network of relationships, a purely connectional system, in which all of the information is in its connectivity. (Lamb 1998: 65)

This is also exemplified in figure 2.3 above. As can be seen, the symbols '*go*' and 'went' are written at the sides of the connecting lines and are not part of the network structure. In Lamb's view, an integration of these symbols would be superfluous, since all the information about the form *go* is already given in the connectivity of the network, namely that it has a particular meaning (here represented by GO), that it belongs to the syntactic class of verbs, and that it is realized by the three phonemes on the left of the bottom line of the figure. That is, the information about the form *go* lies in the connection of these particular parts of the network. Similarly, *went* is superfluous in this network, since this form is represented by all those parts of the network that lead up from the four phonemes on the right-hand side of the bottom line in figure 2.3.

At the basis of Richard Hudson's (1984, 1990, 2007a, 2007b, 2010) Word Grammar is the Network Postulate which states that "language is a conceptual network" (2007a: 1; see chapter 1). Figure 2.4 illustrates the main components of Hudson’s model. The lines with the triangular base signify ISA-relations. As in the model by Collins and Quillian, these relations are fundamental, since they guarantee that properties of a category are inherited by every member of that category and by every member of every sub-category (unless they are overridden). In the figure above b ISA a, which means that b inherits all features from a. The arrows in the model relate two nodes in such a way that the node at the endpoint of the arrow is the function of the starting point of the arrow, i.e. “the e of a is c”, e.g. the/a property of a bird is that it can sing. Hudson claims "that this notation applies throughout language, from phonology through morphology and syntax to semantics and sociolinguistics" (2007b: 511). The network model to be advocated in this study is similar to Hudson's model in that it tries to account for the vast range of linguistic phenomena by a fairly simple notational apparatus.

Hudson follows Lamb (1998) when he claims that "the nodes are defined *only* by their links to other nodes; [...] No two nodes have exactly the same links to exactly the same range of other nodes, because if they did they would by definition be the same node” (Hudson 2007b: 520). That is, all the information is contained in the network itself, labels are a mere representational device and, therefore, redundant.

As can be seen from figure 2.4, some of the links in the network model are labelled. These labels are essential, because Hudson's network is not a mere associative network, as is the one suggested by Collins and Loftus (1975: 412) (see figure 2.2). However, just like nodes, links are also organized in a network

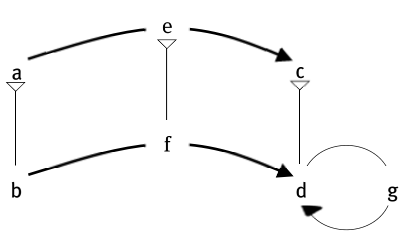

Figure 2.4: Notation in Hudson’s Word Grammar (2007b: 512) (with kind permission of Richard Hudson).

of hierarchical classification, which means that they can simply be identified by their relation to other links. Again, the label is a mere representational device. We will leave it at that for the moment and discuss Hudson's model at greater length at different points in this chapter when we contrast it to the network model developed in this study.

All of the models so far are what is called ‘hard-wired’, i.e. it is the researcher who determines which nodes in the network should be connected to each other and how strong the connections should be. In this way, it is possible to model any aspect of a given language despite such networks not being able to learn. A cognitively plausible model, in addition to dispensing with symbols (as Lamb’s and Hudson’s models do), should also be self-organizing, as MacWhinney (2000: 123) makes clear. However, “[w]hen the prohibition against symbol passing is combined with the demand for self-organization, the class of potential models of language learning becomes extremely limited” (MacWhinney 2000: 124). As a consequence such models are usually confined to highly restricted aspects of the language system only, such as past tense morphology (Rumelhart and McClelland 1986; MacWhinney and Leinbach 1991, Plunkett and Marchman 1993), spelling (Bullinaria 1997), reading (Bullinaria 1997) or acquisition of syntactic categories like *NOUN* and *ADJECTIVE* (Gasser and Smith 1998).

Without going into too much detail, many models of the self-organizing kind all have a similar architecture consisting of a layer of input nodes[8], a layer of output nodes and one or more intermediate layers of hidden nodes. Let us consider an example of a network that ‘learns’ the assignment of the correct form of the direct article to German nouns (MacWhinney et al. 1989). The network is shown in figure 2.5 below.

As can be seen from this figure, the network consists of a total of 66 (=11+5+15+16+2+17) input nodes (we will discuss presently what these nodes represent). Each of these input nodes is connected to a first layer of hidden nodes. More specifically, 49 of the input nodes are connected to 20 gender/number nodes and 19 of the input nodes are connected to 10 case nodes. The 30 nodes in first layer of hidden nodes are connected to 7 nodes on a second hidden layer. These are connected to 6 output nodes, which represent the six possible forms of the German definite article. The layers of hidden nodes can be understood as representing those parts of the language system that contain the knowledge relevant for the choice of the definite article; MacWhinney et al. (1989) write: “We can think of these internal layers as forming a useful internal

8 I will use the term ‘node’ instead of ‘unit’ here.