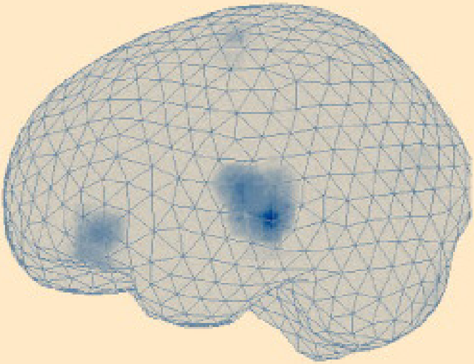

The Neuroscience of Language

On Brain Circuits of Words and Serial Order

Friedemann Pulvermüller

CAMBRIDGE

CAMBRIDGE

more information - www.cambridge.org/9780521790260

The Neuroscience of Language

On Brain Circuits of Words and Serial Order

A realistic model of language should specify the mechanisms underlying language use and comprehension. A neurobiological approach has been shown to be an effective means toward this end. *The Neuroscience of Language* provides results of brain activation studies, patients with brain lesions, and hints from computer simulations of neural networks to help answer the question: How is language organized in the human brain?

At the book's core are neuronal mechanisms that is, the nerve cell wiring of language in the brain. Neuronal models of word and serial-order processing are presented in the form of a computational and connectionist neural network. The linguistic emphasis is on words and elementary syntactic rules. The book introduces basic knowledge from disciplines relevant in the cognitive neuroscience of language. Introductory chapters focus on neuronal structure and function, cognitive brain processes, the basics of classical aphasia research and modern neuroimaging of language, neural network approaches to language, and the basics of syntactic theories. The essence of the work is contained in chapters on neural algorithms and networks, basic syntax, serial-order mechanisms, and neuronal grammar. Throughout, excursuses illustrate the functioning of brain models of language, some of which are simulations accessible as animations on the book's accompanying web site.

This self-contained text and reference puts forth the first systematic model of language at a neuronal level that is attractive to language theorists but that is also well grounded in empirical research. *The Neuroscience of Language* bridges the gap between linguistics and brain science, appealing to advanced students and researchers in neuroscience, linguistics, and computational modeling.

Friedemann Pulvermüller is a senior scientist at the Medical Research Council Cognition and Brain Sciences Unit in Cambridge.

The Neuroscience of Language

On Brain Circuits of Words and Serial Order

FRIEDEMANN PULVERMÜLLER
Cognition and Brain Sciences Unit
Medical Research Council

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo

Cambridge University Press
The Edinburgh Building, Cambridge CB2 2RU, United Kingdom

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org

Information on this title: www.cambridge.org/9780521790260

First published in print format 2003

ISBN-13 978-0-511-06921-5 eBook (EBL)
ISBN-10 0-511-06921-9 eBook (EBL)

ISBN-13 978-0-521-79026-0 hardback
ISBN-10 0-521-79026-3 hardback

ISBN-13 978-0-521-79374-2 paperback
ISBN-10 0-521-79374-2 paperback

Contents

Preface

How is language organized in the human brain? This book provides results of brain activation studies, facts from patients with brain lesions, and hints from computer simulations of neural networks that help answer this question. Great effort was spent to spell out the putative neurobiological basis of words and sentences in terms of nerve cells, or *neurons*. The *neuronal mechanisms* – that is, the nerve cell wiring of language in the brain – are actually in the focus. This means that facts about the activation of cortical areas, about the linguistic deficits following brain disease, and the outcome of neural network simulations will always be related to neuronal circuits that could explain them, or, at least, could be their concrete organic counterpart in the brain.

In cognitive neuroscience, the following questions are commonly asked with regard to various higher brain functions, or *cognitive processes*, including language processes:

(1) Where? In which areas of the brain is a particular process located?
(2) When? Before and after which other processes does the particular process occur?
(3) How? By which neuron circuit or which neuron network type is the particular process realized?
(4) Why? On the basis of which biological or other principles is the particular process realized by this particular network, at this particular point in time, and at these particular brain loci?

The ultimate answer to the question of language and the brain implies answers to these questions, with respect to all aspects of language processing. The aspects of language relevant here include the physical properties of speech sounds and the sound structure of individual languages as specified by phonetics and phonology, the meaning and use of words and larger units of language as specified by semantics and pragmatics, and the rules underlying the serial ordering of meaningful language units in sentences or larger sequences as specified by syntax. All of these aspects are addressed, although an emphasis is put on words and elementary syntactic rules.

Lesion studies and brain imaging studies have revealed much important information relevant for answering "Where" questions of type (1), about the relevant brain loci. Fast neurophysiological imaging could also shed light on the temporal structure of language processes in the millisecond range, thereby answering "When" questions of type (2). Type (3) "How" questions about the underlying *mechanisms* are sometimes addressed in neural network studies, but many ideas – in particular, ideas about syntactic structures – are still formulated in terms so abstract that it is difficult to see possible connections to the neuronal substrate of the postulated processes. A common excuse is that we do not know enough about the brain to specify the mechanisms of grammar, and of language in general, in terms of neurons. I never found this to be a very good excuse. Only if we theorize about concrete language circuits can we ever understand them. Therefore, this book puts forward concrete, or *neuronal*, models of word and serial-order processing. The introduced models are linked to general neuroscientific principles, and are thereby used to answer aspects of "Why" questions of type (4) as well.

This book came about because my abstract theorizing about neuron circuits that could realize aspects of language was initiated by various discussions with Valentino Braitenberg and Almut Schüz, with whom I had the pleasure of collaborating at the Max Planck Institute of Biological Cybernetics in Tübingen, Germany. This was before and at the very start of the so-called Decade of the Brain, with much less data in hand than is available today. However, when collecting brain imaging data myself in subsequent years, I found it necessary to have a brain-based model of language as a guideline for designing experiments. I also found the concepts and mechanisms developed and envisaged earlier quite helpful for the interpretation and explanation of new results. In the context of neuroimaging studies, I should mention Werner Lutzenberger, my most important teacher in the neurophysiological domain. I thank Almut, Valentino, and Werner for the countless discussions about theory and data that laid the foundation for this book.

However, many more colleagues and friends contributed substantially. Robert Miller was a critical reader of many manuscripts preceding this text and provided me with indispensable knowledge, particularly neuroanatomical in nature. Stefano Crespi-Reghizzi, an expert in automata theory and computational approaches to language, was kind enough to teach me the basics from these disciplines, thereby preventing me from serious mistakes.

Helmut Schnelle, a theoretical linguist and one of the few who seriously attempt at connecting syntax and the brain, provided me with a very detailed and equally helpful critique of an earlier version of this book. Comparison with his approach to the representation of serial order in the brain helped me greatly when formulating putative grammar mechanisms. Detlef Heck checked some of the neurophysiology and neuroanatomy sections and helped me avoid errors there. William Marslen-Wilson read the entire book with the eyes of a psycholinguist, and provided me with probably the most detailed critique of an earlier version of this book, thereby initiating several improvements. I am indebted to them all for their substantial help. I also thank numerous other colleagues who commented on text passages or ideas, including Joe Bogen, Michel Caillieux, Thomas Elbert, Gerd Fritz, Joaquin Fuster, Sarah Hawkins, Risto Ilmoniemi, Risto Näätänen, Dennis Norris, Lee Osterhout, Günther Palm, Brigitte Rockstroh, Arnold Scheibel, John Schumann, and, of course, my closest colleagues, Ramin Assadollahi, Olaf Hauk, Bettina Neininger, Yury Shtyrov, and, most importantly, my wife Bettina Mohr, not only because she performed the first experiment that made me feel that the kind of model put forward here might be on the right track, but also because she continuously gave me the radical and respectless critique that is a necessary condition for scientific progress. Finally, I apologize to my son Johannes David for the untold goodnight stories and unplayed games that fell victim to writing the present pages.

Friedemann Pulvermüller
Cambridge, July 2001

The Neuroscience of Language

On Brain Circuits of Words and Serial Order

CHAPTER ONE

A Guide to the Book

The neuroscience of language is a multidisciplinary field. The reader's primary interest may therefore lie in various classical disciplines, including psychology, neuroscience, neurology, linguistics, computational modeling, or even philosophy. Because readers with different backgrounds may be interested in different parts of this book, Chapter 1, Section 1.3 gives an overview of the book contents and the gist of each chapter. In Section 1.1, the general structure of the book is explained; subsequently, paths through the book are recommended for readers with different backgrounds and interests in Section 1.2.

1.1 Structure and Function of the Book

The fourteen chapters of this book are mainly designed to convey one single message: It is a good idea to think about language in terms of brain mechanisms – to spell out language in the language of neurons, so to speak. Making this point is not a new proposal. One can find similar statements in classical writings; for example, in Freud's monograph on aphasia (Freud, 1891) and other publications by neurologists in the late nineteenth century, and, of course, in modern brain-theoretical and linguistic publications (Braitenberg, 1980; Mesulam, 1990; Schnelle, 1996a). However, a systematic model of language at the level of neurons as to date is not available, at least, not an approach that would be both grounded in empirical research while at the same time attacking a wide range of complex linguistic phenomena.

Apart from the main message, this book puts forward two principle proposals: First, that words are represented and processed in the brain by strongly connected distributed neuron populations exhibiting specific topographies. These neuron ensembles are called *word webs*. Second, that grammar mechanisms in the brain can be thought of in terms of neuronal assemblies whose activity specifically relates to the serial activation of pairs of other neuron ensembles. These assemblies are called *sequence sets*. The proposal about word webs is presented in Chapter 4 and the one about sequence sets in Chapter 10. One may therefore consider Chapters 4 and 10 the core chapters of this book.

As it happens, new proposals elicit discussion, which, in turn, makes refinement of the original proposals desirable. The word web proposal is being refined in Chapters 5 and 8, and the proposal on grammar mechanisms is further developed in Chapters 11 and 12. As stressed in the Preface, several colleagues contributed to the refinements offered. The evolution of some of the ideas is documented in a recent discussion in the journal *The Behavioral and Brain Sciences* (Pulvermüller, 1999b). Summaries of ideas put forward here can be found in related review papers (Pulvermüller, 2001, 2002).

Apart from presenting the two main proposals, the book is designed to give the reader an introduction to basic knowledge from disciplines relevant in the cognitive neuroscience of language. Chapter 2 offers an introduction to neuroscience and cognitive brain processes. Chapter 3 introduces basics about classical aphasia research and modern neuroimaging of language. Two more introductory chapters follow approximately in the middle of the book. Chapter 6 features neural network approaches to language, and Chapter 7 introduces basics of syntactic theories. These introductory chapters were written to make the book "self-contained," so that ideally speaking no prior special knowledge would be required to understand it.

Interspersed between the chapters are five excursuses, labeled E1 through E5, which illustrate the functioning of brain models of language. In each excursus, one or more simple simulations are summarized that address an issue raised in the preceding chapter. Computer simulations of the main syndromes of aphasia (Excursus E1) are included along with simulations of the processing of simple (Excursus E2) and gradually more complex (Excursuses E3–E5) sentences in brain models of grammar. Some of the simulations are available as animations accessable through the Internet.

1.2 Paths Through the Book

Clearly, the reader can choose to read through the book from beginning to end. However, because not all issues covered by the book may be in the inner circle of one's personal "hot topics," it may be advantageous to have available alternatives to this global strategy. One alternative would be to take a glance at the main chapters (4 and 10) or at the introductory chapter concerning the topic one is particularly keen on. However, one may

Table 1.1. Routes through the book recommended to readers primarily interested in neuroscience, linguistics, or neuronal modeling, respectively. Chapter numbers and headings are indicated. Headings are sometimes abbreviated. Excursuses are referred to by the letter E plus a number and by abbreviated headings. For further explanation, see text.

Neuroscience Route	Linguistics Route	Modeling Route
2 Neuronal structure and function	4 Words in the brain	4 Words in the brain
3 Aphasia and neuroimaging	7 Basic syntax	5 Regulation, overlap, web tails
4 Words in the brain	8 Synfire chains	6 Neural networks
E1 Double dissociations	9 Sequence detectors	E1 Double dissociations
5 Regulation, overlap, web tails	10 Neuronal grammar	E2 Basic bits of neuronal grammar
8 Synfire chains	11 Neuronal grammar and algorithms	E3 Web response to a sentence
9 Sequence detectors	12 Refining neuronal grammar	E4 Lexical ambiguity
13 Neurophysiology of syntax	14 Linguistics and the brain	E5 Center embedding

wish to dive deeper into the matter while still primarily following one's interests.

For this latter purpose, three paths through the book are offered for a reader primarily interested in neuroscience, linguistics, and neurocomputational modeling. If one chooses one of these options, one should be aware that the routes are not self-contained and consultation of other chapters may be relevant occasionally. To facilitate detection of relevant information in other chapters of the book, multiple cross-references have been added throughout.

The three paths through the book are presented in Table 1.1. Please consult the overview, Section 1.3, for details about chapter contents.

It is difficult to decide what to recommend to a reader primarily interested in psychology. Because psychology is a rather wide field, the best recommendation may depend primarily on the subdiscipline of interest. Readers interested in neuropsychology and psychophysiology can be recommended to follow the neuroscience route, whereas those interested in cognitive psychology may tend more toward modeling aspects. The neuroscience route would also be recommended to the reader focusing on neuroimaging or neurology. A philosopher may be most interested in the open questions that accumulate in Chapters 5, 12, and 14.

1.3 Chapter Overview

1.3.1 Chapter 1: A Guide to the Book

The main purpose of the book and its structure are explained briefly. Recommendations are given concerning how to use the book if one is interested primarily in its neuroscience, linguistics, or modeling aspects. The gist of each book chapter is summarized briefly.

1.3.2 Chapter 2: Neuronal Structure and Function

Chapter 2 introduces basics about the anatomy and function of the neuron and the cortex. Principles of cortical structure and function are proposed that may be used as a guideline in cognitive brain research. The concept of a distributed functional system of nerve cells, called *functional web*, is introduced and discussed in the light of neurophysiological evidence.

1.3.3 Chapter 3: From Aphasia Research to Neuroimaging

Basics about *aphasias*, language disorders caused by disease of the adult brain, are summarized. Aphasia types and possibilities on explaining some of their aspects are being discussed. The issue of the laterality of language to the dominant hemisphere – usually the left hemisphere – is mentioned, and theories of laterality and interhemispheric interaction are covered. Basic insights in the functional architecture of the cortex as revealed by modern neuroimaging techniques are also in the focus. The conclusion is that some, but not all, insights from classical aphasia research about the localization of cortical language functions can be confirmed by neuroimaging research. However, language processes seem to be much more widely distributed than previously assumed. The question about the cortical locus of word semantics, as such, has found contradicting answers in recent imaging research.

1.3.4 Chapter 4: Words in the Brain

The proposal that words are cortically represented and processed by distributed functional webs of neurons is elaborated and discussed on the basis of recent neuroimaging studies. The data support the postulate that words and concepts are laid down cortically as distributed neuron webs with different topographies. The strongly connected distributed neuron ensembles representing words are labeled *word webs*. Word webs may consist of a phonological part (mainly housed in the language areas) and a semantic part (involving other areas as well). For example, processing of words with strong associations to actions and that of words with strong visual associations appears to activate distinct sets of brain areas. Also, different subcategories of action words have been found to elicit differential brain responses. This supports the proposed model.

1.3.5 Excursus E1: Explaining Neuropsychological Double Dissociations

A simulation is presented that allows for the explanation of neuropsychological double dissociations on the basis of distributed functional webs of neurons. The nonlinear decline of performance of the models with lesion size and its putative neurological relevance are also mentioned.

1.3.6 Chapter 5: Regulation, Overlap, and Web Tails

Chapter 5 deals with open issues remaining from earlier chapters. How could a regulation device controlling activity in the cortex be organized? How would words with similar meaning but different form, or words with similar form but different meaning, be realized in the brain? Would the brain's word processor be restricted to the cortex, or can word webs have subcortical "tails"? One postulate is that multiple overlap between cortical representations exists between word representations.

1.3.7 Chapter 6: Neural Algorithms and Neural Networks

An introduction into neural network models is given. McCulloch and Pitt's theory is sketched and perceptron-based simulations are featured. Symbolic connectionist approaches are also discussed briefly. Among the hot topics featured are the explanation of word category deficits as seen in neurological patients and the development of rules in infants' brains.

1.3.8 Chapter 7: Basic Syntax

A few terms and theoretical approaches to syntax are introduced. Phrase structure grammars, dependency grammars, and more modern proposals rooted in these classic approaches are discussed. Syntactic problems such as those associated with long-distance dependencies and center embeddings are mentioned. Chapter 7 ends with a list of issues with which grammar circuits should cope.

1.3.9 Chapter 8: Synfire Chains as the Basis of Serial Order in the Brain

One type of serial-order mechanism in the brain for which there is evidence from neurophysiological research is featured. Called a *synfire chain*, it consists of local groups of cortical neurons connected in sequence, with loops also allowed for (*reverberatory synfire chain*). The synfire model of serial order is found to be useful in modeling phonological–phonetic processes. It is argued, however, that a synfire model of syntax does not appear to be fruitful.

1.3.10 Chapter 9: Sequence Detectors

A second type of serial-order mechanism exists for which there is evidence from brain research. It is the detection of a sequence of neuron activations by a third neuronal element called the *sequence detector*. The evidence for sequence detectors comes from various brain structures in various creatures. It is argued that sequence detectors may operate on sequences of activations of word webs, and that these may be part of the grammar machinery in the brain.

1.3.11 Chapter 10: Neuronal Grammar

Neuronal sets are defined as functional webs with four possible activity states: inactivity (O), full activation or ignition (I), sustained activity or reverberation (R) and neighbor-induced preactivity or priming (P). Reverberation and priming levels can vary. Grammar networks are proposed to be made up of two types of neuronal sets: word webs and sequence sets. Sequence sets respond specifically to word sequences. The lexical category of words and morphemes is represented by a set of sequence sets connected directly to word webs. Words that can be classified as members of different lexical categories have several mutually exclusive sets of sequence sets. Activity dynamics in the network are defined by a set of principles. A grammar network, also called *neuronal grammar*, can accept strings of words or morphemes occurring in the input, including sentences with long-distance dependencies. The hierarchical relationship between sentence parts becomes visible in the activation and deactivation sequence caused by an input string.

1.3.12 Chapter 11: Neuronal Grammar and Algorithms

Three types of formulas are introduced that describe a neuronal grammar network:

1. *Assignment formulas* are definitions of connections between input units and lexical category representations and are analogous to lexicon or assignment rules of traditional grammars.
2. *Valence formulas* are definitions of lexical categories in terms of sequencing units and have some similarity to dependency rules included in dependency grammars.
3. *Sequence formulas* are definitions of connections between sequencing units and have no obvious counterpart in traditional grammars.

1.3.13 Excursus E2: Basic Bits of Neuronal Grammar

Simple word strings are discussed on the basis of grammar networks composed of sequence sets and word webs. How the network accepts a string and how the network behaves if it fails to do so is discussed.

1.3.14 Excursus E3: A Web Response to a Sentence

Processing of an ordinary sentence is simulated in a neuronal grammar architecture. The sentence exhibits six morphemes, subject–verb agreement, a distributed word, and other interesting properties.

1.3.15 Chapter 12: Refining Neuronal Grammar

A revision of the grammar model is proposed that requires stronger assumptions. The core assumption is that neuronal sets exhibit multiple states of reverberation and priming. In the new architecture, the relationship between words and lexical categories is now dynamic.

1.3.16 Excursus E4: Multiple Reverberation for Resolving Lexical Ambiguity

Implementation of multiple lexical category representations of words using mutually exclusive sets of sequence sets allows for modeling sentences in which the same word form is being used twice, as a member of different lexical categories.

1.3.17 Excursus E5: Multiple Reverberations and Multiple Center Embeddings

A network with dynamic binding between word and lexical category representations and the option to activate each neuronal set is introduced on the background of the machinery discussed in Chapters 10 and 11. This more advanced architecture now models the processing of grammatically complex sentences that include center embeddings.

1.3.18 Chapter 13: Neurophysiology of Syntax

Grammatically incorrect "sentences" elicit specific physiological brain responses. Two such physiological indicators of grammatical deviance are discussed. The neuronal grammar proposal is related to these data, and a putative neurobiological explanation for them is offered.

1.3.19 Chapter 14: Linguistics and the Brain

Linguistics and brain science must merge. This is reemphasized in Chapter 14 where putative advantages of a brain-based language theory are highlighted.

CHAPTER TWO

Neuronal Structure and Function

A realistic model of language must specify the mechanisms underlying language use and comprehension. What are the relevant mechanisms? It is certain that it is the human brain that provides the mechanisms realizing language, and it is almost equally certain that language mechanisms are organized as nerve cells and their mutual connections. A *realistic* model of language, therefore, must specify the putative organic basis of language use and language comprehension in terms of neurons, neuronal connections, and neuron circuits. This does not necessarily mean that the model must specify each and every single neuron that participates, but it does mean that the circuits believed to underlie language function should be specified as far as possible and relevant. Rather than saying that a language sound, word, or syntactic rule is represented in the brain, period, one may wish to learn in which way such sounds, words, or rules are laid down. Therefore, it is necessary to introduce known neuronal mechanisms and others that can be inferred from more recent data.

Chapter 2 gives a brief introduction to neuronal mechanisms, with special emphasis on those mechanisms that may be relevant for organizing language. The chapter first addresses questions about neuronal architecture or structure (Section 2.1). What is a nerve cell or neuron? What is the global structure of the *cortex*, the portion of the brain most important for language? The idea here is that these anatomic structures are related to the computations in which they are involved. The brain machinery is not just one arbitrary way of implementing the processes it realizes, as, for example, any hardware computer configuration can realize almost any computer program or piece of software. The claim is that, instead, the hardware reveals aspects of the program. *Neuronal structure is information* (Braitenberg, 1971). In other words, it may be that the neuronal structures themselves teach us about aspects of the computational processes that are laid down in these structures.

The paragraphs on cortical structure detail a few structural features and elaborate on their functions as well.

As a next step, additional functional properties of the neuron are highlighted – in particular, the question of how information may be stored in nerve cells and their connections is addressed (Section 2.2). The structural and functional properties of the neuron and the cortex are summarized by three conclusions that are used later in the book as axioms or principles for theorizing about language mechanisms (Section 2.3). A few thoughts follow about how neurons in the cortex may interact to yield what is sometimes called the *cognitive* or *higher* brain functions (Section 2.4). These terms can refer to language-related processes, of course, but to other complex perceptual and action-related processes as well. Finally, the concept of a *functional web* is introduced and grounded in empirical evidence.

2.1 Neuronal Structure

2.1.1 Anatomy of a Nerve Cell

The functional element of the brain and the nervous system is the nerve cell, or *neuron*. Figure 2.1 shows an example of one type of neuron, so-called *pyramidal cells*. This is the most common neuron type in the largest structure of the human brain, the cortex. Like most other neuron types, a pyramidal cell consists of *dendrites*, a *cell body*, and an *axon*. In Figure 2.1, the cell body is the thick speck in the middle. The thicker lines departing from the cell body and running sideways and upward are the dendrites, and the thin line running downward is the axon. The dendrites and the axon branch multiply. The neuron receives signals from other neurons through its dendrites, and it transmits its own signals to other neurons through its axon and its many branches. Whereas the dendrites are short and hardly reach loci 1 mm away from the cell body, many pyramidal neurons have very long axon branches, in addition to the short axon branches shown in the figure. The long axon branch of a neuron can be 10 cm or more long, and its synapses can contact other neurons in distant cortical areas or subcortical nuclei.

Signals are passed between neurons by contact buttons called *synapses*. The synapses delivering incoming – or *afferent* – signals to the neuron are located mainly on the dendrites and a few sit on the cell body. The dendrites of pyramidal cells appear to be relatively thick in Figure 2.1 because they carry many hillocks, called *spines*, on which most of the afferent synapses are located (Fig. 2.2). The spine's size and shape may be related to the functional connection between two pyramidal neurons. Outgoing – or *efferent* – signals are transmitted through synapses at the very ends of the axon's branches.

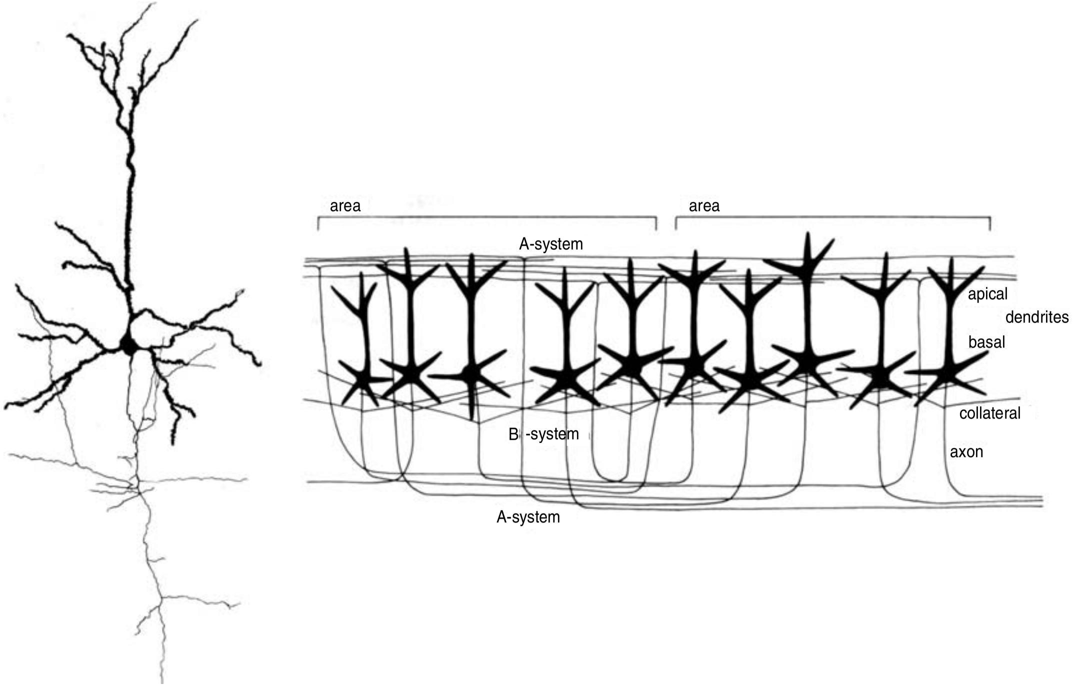

Figure 2.1. (Left) Golgi-stained pyramidal cell. The cell body is in the middle and the apical and basal dendrites can be distinguished along with the axon and its local branches. The long axon branch running into the white matter and connecting with neurons in other brain areas is not shown. Reprinted with permission from Braitenberg, V., & Schüz, A. (1998). *Cortex: statistics and geometry of neuronal connectivity* (2 ed.). Berlin: Springer (Right) The cortical wiring through the long-distance A-system and local B-system of connections is illustrated. (For explanation, see text.) From Braitenberg, V. (1991). Logics at different levels in the brain. In I. Cluer (Ed.), *Atti del Congresso "Nuovi problemi della logica e della filosofia della sienza," Vol. 1* (pp. 53–9). Bologna: Franco Angeli Editore.

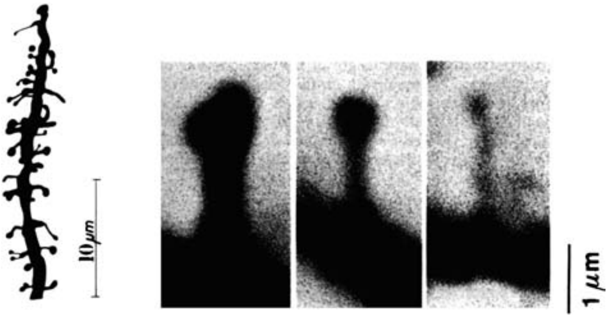

Figure 2.2. (Left) A dendrite with spines covering its surface. (Right) Microphotographs of spines with different calibers. Most synapses between pyramidal cells are located on spines. Their shape influences the efficacy of signal transmission between the cells. From Braitenberg, V., & Schüz, A. (1998). *Cortex: statistics and geometry of neuronal connectivity* (2 ed.). Berlin: Springer.

A pyramidal neuron in the human cortex has about 5 × 10,000 (10^4) incoming synapses and about the same number of efferent synapses (DeFelipe & Farinas, 1992).

As mentioned, the dendrites of pyramidal cells branch multiply, resulting in a so-called dendritic tree. This treelike structure differs from the shape of a typical conifer. It has sideways branches, called *basal dendrites*, and in many cases one long upward branch, called the *apical dendrite*. This anatomical subdivision in apical and basal dendrites has functional meaning (Braitenberg, 1978b; Braitenberg & Schüz, 1998). Neurons located in close vicinity appear to contact each other primarily through synapses on their basal dendrites. Apical dendrites in upper cortical layers carry synapses contacting distant neurons located in subcortical structures or in other areas of the cortex. A similar subdivision has been proposed for the axon's branches. Some branches are short and contact neighboring neurons. However, the long axon reaching distant cortical areas has many additional branches at its end and contacts other neurons there, usually on their apical dendrites.

Thus, already the microanatomy of the pyramidal neuron suggests a subdivision of cortical connections into a local and a long-distance system. The local system is between basal dendrites and local axon collaterals. Because it is wired through basal dendrites, it has been dubbed the *B-system*. The long-distance system is between long axons and the apical dendrites they contact and is therefore called the *A-system*. About half of a pyramidal neuron's synapses are in the short-distance B-system, and the other half in the long-range A-system. It has been speculated that the local B-system and the wide-range A-system have different functions and play different roles in cognitive processing (Braitenberg, 1978b).

2.1.2 Basics of the Cortex

The part of the brain most relevant for language is the cerebral cortex. This fact has been proved by neurological observations, in particular, the fact that lesions in certain areas of the cortex lead to neurological language impairment, aphasia, and by more recent imaging studies. The cortex consists of two halves called the cortical hemispheres, one of which is usually more important for language than the other (Broca, 1861).

The cortex is an amply folded layer and, therefore, much of its surface is not visible when looking at it from outside, but rather buried in the many valleys or *sulci*. The cortical parts visible from the outside are the *gyri*. In humans, the cortex is only 2- to 5-mm thick. This thin layer of densely packed neurons rests on a fundament of white matter. The white matter has its color from the isolating sheaths covering most of the A-system axons of cortical cells. Compared to the shining white of the white matter, the cortex itself looks gray because of the many gray cell bodies it contains. The white matter fundament of the cortex is much more voluminous than the gray matter. This mere anatomical fact – the thin gray matter layer of cell bodies sitting on a massive fundament of cables connecting the cell bodies – suggests that the information exchange between cortical neurons that lie far apart may be an important function of this structure.

The human cortex includes 10^{10} to 10^{11} neurons, about 85 percent of which are pyramidal cells (Braitenberg & Schüz, 1998). Because each of these neurons has about 10^4 synapses, the total number of cortical synapses is in the order of 10^{14} or even higher. Apart from pyramidal cells, there are other cell types. Some of these have inhibitory function. These neurons are much smaller than pyramidal cells, and make only local connections and receive input from many adjacent pyramidal cells. The small inhibitory cells can dampen down cortical activity in case too many of their adjacent excitatory pyramidal cells become active at one time.

Each hemisphere can be subdivided into some 50 to 100 areas. Figure 2.3 shows a lateral view of the left hemisphere with its most common area subdivision. These areas have been proposed by Brodmann (1909) and reflect neuroanatomical properties of the cortical gray matter.

A cruder subdivision of the cortex is in terms of lobes. Figure 2.3 shows the frontal lobe on its upper left, the parietal lobe at the top, the occipital lobe on the right, and the temporal lobe at the bottom. Two major landmarks indicate

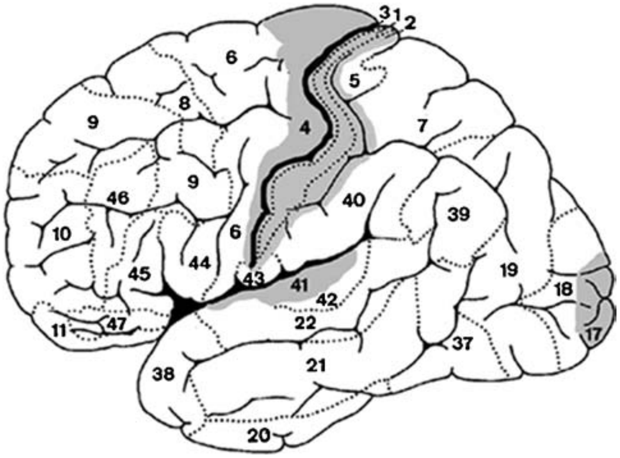

Figure 2.3. Brodmann's map of the human cortex. A lateral view on the left hemisphere is shown and the area numbers are indicated. Solid lines indicate area numbers and broken lines indicate boundaries between areas. The shaded fields are primary areas. From Brodmann, K. (1909). *Vergleichende Lokalisationslehre der Großhirnrinde*. Leipzig: Barth.

boundaries between lobes. The *central sulcus*, or fissure, indicated by the bold line between Brodmann areas 3 and 4, is the boundary between the frontal and the parietal lobes, and the *Sylvian fissure*, running horizontally below areas 44 and 45 and above area 22, separates the temporal lobe from both frontal and parietal lobes. The Sylvian fissure is important as a landmark, because all of the areas most relevant for language are located in its close vicinity. The areas next to the Sylvian fissure are called the *perisylvian areas*. The occipital lobe at the back of the cortex is less easy to define based on gross anatomical landmarks. Gyri and sulci are sometimes referred to by specifying their locus within a lobe. In the temporal lobe, for example, an upper (superior), a middle, and a lower (inferior) gyrus can be distinguished, and the same is possible in the frontal lobe, where, for example, the inferior gyrus includes Brodmann areas 44 and 45.

When voluntary movements or actions are being performed, the necessary muscle contractions are caused by activity in the cortex. When stimuli cause perceptual experiences, there is also activity in the cortex. The efferent fibers through which the cortex controls muscle activity and the afferent fibers transmitting information from the sensory organs to the cortex originate from defined areas. These are the *primary* and *secondary areas*. The primary areas most relevant for language processing are as follows:

- Brodmann area 17 in the posterior occipital lobe, where the fibers of the visual pathway reach the cortex (primary visual cortex)
- Brodmann area 41 in the superior temporal lobe, where the fibers of the auditory pathway reach the cortex (primary auditory cortex)
- Brodmann areas 1–3 in the postcentral gyrus of the parietal lobe, where somatosensory input reaches the cortex (primary somatosensory cortex)
- Brodmann area 4 in the precentral gyrus of the frontal lobes, whose efferent neurons contact motor neurons and control muscle contractions (primary motor cortex).

These sensory and motor fields are the shaded areas in Figure 2.3. Also areas adjacent to the primary areas include efferent or afferent cortical connections (e.g., He, Dum, & Strick, 1993). There are additional sensory pathways for olfactory and gustatory input that project to brain areas not shown in the diagram.

Each of the motor and sensory cortical systems is characterized by *topographical* order of projections, meaning that adjacent cells in the sensory organs project to adjacent cortical neurons, and adjacent body muscles are controlled by adjacent neurons in the motor cortex.

The somatotopy of the primary motor cortex is illustrated in Figure 2.4 (Penfield & Roberts, 1959). In this diagram, a frontal section of the cortex along the precentral gyrus is schematized, and the body parts whose movement can be caused by electrical stimulation of the respective part of the cortex is illustrated by the homunculus picture. The representations of the articulators – for example, tongue, mouth, and lips – are adjacent to each other and lie at the very bottom of the illustration. Superior to them, the hand representation is indicated, and the representation of the lower half of the body, including the feet and legs, is further up and in the medial part of the

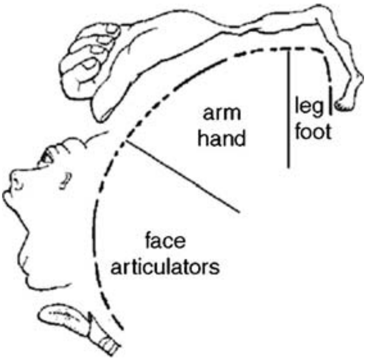

Figure 2.4. Penfield and Rassmussen (1950) investigated the organization of the human motor cortex by stimulating the cortical surface with weak currents and by recording which body muscles became active as a consequence of stimulation. The body muscles whose contraction was caused by stimulation of these areas are indicated by the homunculus picture. The map is consistent with the neuroanatomical projections of the motor cortex. It illustrates the somatotopic organization of the primary motor cortex. From Penfield, W., & Rassmussen, T. (1950). *The cerebral cortex of man*. New York: Macmillan.

cortex. There are topographically ordered projections between body muscles and loci of the primary motor cortex, and topographically ordered projections exist in the visual, auditory, and somatosensory cortical systems as well.

These projections are established early in life. However, they can be altered, for example, as a consequence of brain injuries. Research on cortical reorganization has shown that the cortical representations can change dramatically, for example, as a consequence of deprivation (Buonomano & Merzenich, 1998; Kujala, Alho, & Näätänen, 2000; Merzenich et al., 1983). A change of the cortical areas involved in processing a particular input can even be a consequence of learning. For example, string players and braille readers show an altered function of somatosensory areas with enlarged cortical representations of the extremities involved in complex sensorimotor skills (Elbert et al., 1995; Sterr et al., 1998). This shows that the topographic projections are not fixed genetically, but may vary within certain boundaries. Because sensory projections to the cortex are being altered following deprivation and the acquisition of special skills, the projection map in Figure 2.4 is probably appropriate only for nondeprived individuals without special skills. Nevertheless, even after substantial cortical reorganization, the principle of topographical projections still appears to hold true (Merzenich et al., 1983).

2.1.3 Internal Wiring of the Cortex

Apart from the connections between the cortex and the outside world, it is relevant to say a few words about the internal wiring of the cortex. The anatomy of the pyramidal cell makes it clear that there are different types of connections: short-range and long-range. As mentioned, the local links are through the B-system, that is, through connections between basal dendrites and axon collaterals. The short-range connections are between neighboring neurons whose trees of dendrites and local axon collaterals overlap. The likelihood of such a contact between two neurons occupying almost the same cortical space appears to be high. Estimates of the connection probabilities p of nearby neurons range between 10 percent and 80 percent, depending on the method used (Braitenberg & Schüz, 1998; Hellwig, 2000; Mason, Nicoll, & Stratford, 1991). Assuming a maximum of two connections, one in each direction, between any two neurons a group of n neurons can have a maximum of $n(n-1)$ direct connections between its members. The estimated number of connections between n nearby neurons, located, say, under the same $0.5\ \mathrm{mm}^2$ of cortical surface, would therefore be $p \times n(n-1)$. This means that 1,000 neurons selected by chance from a small piece of cortex would have hundreds of thousands of mutual synaptic links – the exact number being in the order of 100,000 to 800,000, depending on the actual parameters chosen. This implies that these neurons can exert a strong influence on each other, given they activate each other near synchronously. In essence, the high probability of local B-system links suggests local clusters of neurons that exchange information intensely and therefore show similar functional characteristics. Such local clusters, called columns, have been identified, for example, in the visual system. A column includes neurons below 0.1–0.5 mm^2 of cortical surface. The neurons in a column respond to similar visual stimulus features and can thus be considered to be functionally related (Hubel, 1995). Although this functional relatedness is, in part, due to their common sensory input, within-column connections are also likely to contribute.

The long-distance A-system connections between pyramidal cells link neurons in different cortical areas through their long axons and apical dendrites. However, not every area is connected to every other area. The connection structure of the human cortex may be inferred from the results of neuroanatomical studies in animals, usually cats and macaque monkeys. These animal studies provide evidence that most primary cortical areas do not contact each other through direct connections (Pandya & Yeterian, 1985). The primary motor and somatosensory cortices, which lie adjacent to each other, represent the only exception. Adjacent areas, as a rule, are connected with very high probability ($>$ 70 percent; Young, Scannell, & Burns, 1995). For pairs of distant areas, that is, areas with more than one other area between them, this probability is lower (15–30 percent). Still, however, it is remarkable that, for example, in the macaque monkey, where nearly seventy areas were distinguished, most areas would have links to ten or more distant areas within the same cortical hemisphere. Even in the brain of the mouse, in which only twelve local compartments were distinguished, each compartment was found to send out to and receive projections from five other compartments on average (Braitenberg & Schüz, 1998). In addition, there are connections between most homotopic areas of the two hemispheres. Thus, long-distance links directly connect many, although not all, cortical areas. Another important feature of corticocortical connectivity is that most between-area connections are reciprocal; that is, they include neurons in both areas projecting to the respective other area (Pandya & Yeterian, 1985; Young et al., 1995).

It is not certain that all of these properties of cortical connectivity investigated in monkeys and other higher mammals generalize to humans. A detailed picture of cortical connectivity can be obtained only by using invasive techniques that cannot be applied in humans for ethical reasons, although important insights come from postmortem neuroanatomical studies in humans (Galuske et al., 2000; Jacobs et al., 1993; Scheibel et al., 1985). In particular, conclusions on the pattern of long-distance connections of the areas most important for language must be handled with care because these areas do not have unique homologs in the monkey brain. However, a tentative generalization can be proposed in terms of the position of the areas relative to primary areas.

One such generalization is the following: The primary auditory cortex and the primary motor cortex controlling the articulators are not linked directly. Their connections are indirect, through inferior frontal areas anterior to the mouth–motor cortex and superior temporal areas anterior, posterior, and inferior to the primary auditory cortex (Deacon, 1992; Pandya & Yeterian, 1985). Therefore, if two neurons, one situated in the primary cortex controlling mouth movements and the other in the piece of cortex receiving direct input from fibers of the auditory pathway, want to communicate with each other, they cannot do so directly. They must send their information to relay neurons that link the two areas. Many such relay neurons are likely present in the inferior frontal lobe (Brodmann's areas 44 and 45) and in the superior temporal lobe (Brodmann's areas 22 and 42). A similar point made here for associations between actions and auditory input can also be made for other between-modality associations. Auditory–visual, somatosensory–visual, and visual–action associations all require indirect links through relay neurons in one or more nonprimary areas.

2.2 Neuronal Function and Learning

A neuron receives information, performs calculations on it, and sends out information to other neurons. An afferent synapse on one of its dendrites can produce an activating or inhibiting effect. If a signal is transmitted through the synapse, a so-called postsynaptic potential is elicited. There are *excitatory* and *inhibitory postsynaptic potentials* (EPSPs and IPSPs). The postsynaptic potential is propagated along the dendritic tree toward the cell body. The signal reaching the cell body is crucial for the neuron's output. At the so-called axon hillock, where the axon inserts from the cell body, computations are being performed to the effect that if this signal is excitatory enough, an output is created. If the potential increases rapidly or if it exceeds a threshold value, a condition usually met when several incoming signals arrive simultaneously, a new signal, called the *action potential*, is likely to be elicited and then propagated along the axon. The neuron is said to *fire* when an action potential is being created. The action potential is an all-or-none response – either it is elicited or not, and does not vary in intensity. It propagates without being attenuated from the cell body to the various efferent synapses at the ends of the many branches of the axon. In contrast to the action potential with its all-or-none characteristics, the postsynaptic potential is an analog signal that falls off gradually with time and distance along the dendrite.

Synaptic connections between cortical neurons are usually weak (Abeles, 1991). Several simultaneous or near-simultaneous inputs causing strongly overlapping postsynaptic potentials are therefore necessary to elicit an action potential in a neuron. However, the strength of a synaptic link can vary. Synaptic links become stronger if the synapse is being used frequently. This is the essence of a postulate put forward, among others, by Donald Hebb, who said "that any two cells or systems of cells that are repeatedly active at the same time will tend to become 'associated,' so that activity in one facilitates activity in the other" (1949, p. 70). There is now strong evidence from single- and multiple-unit recordings proving that this postulate is correct (Ahissar et al., 1992; Fuster, 1997; Tsumoto, 1992). If connected neurons fire together, their mutual influence on each other becomes stronger.

Figure 2.5 illustrates the change of synaptic effectiveness caused by coactivation of neurons (Kandel, 1991). A neuron in the hippocampus, a phylogenetically old structure belonging to the cortex, was investigated. If a single input fiber to the neuron was activated for some time, this did not change connection strength. However, when the single input fiber was repeatedly activated together with several others so that the target neuron was caused to fire together with the input, later inputs through the participating synapses had a much stronger effect. This indicates that these synaptic connections

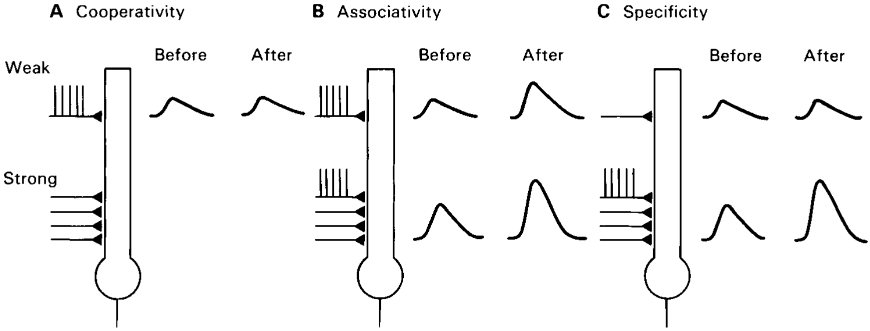

Figure 2.5. Illustration of long-term potentiation. (a) If a single input fiber (uppermost line) to a neuron was activated for some time, this did not change synaptic connection strength. (b) When the single input fiber was activated repeatedly with several others so that the target neuron was caused to fire together with the input, later input through the participating synapses was more effective. (c) Activation of the strong input alone selectively increased the influence of the stimulated neurons, but not that of others. Reprinted with permission from Kandel, E. R. (1991). Cellular mechanisms of learning and the biological basis of individuality. In E. R. Kandel, J. H. Schwartz, & T. M. Jessell (Eds.), *Principles of neural science* (3 ed., pp. 1009–31). New York: Elsevier.

became stronger. The increase of synaptic effectiveness was specific. Activation of the strong input alone did selectively increase the influence of the stimulated neurons, but not that of others. Thus, the strengthening of synaptic links must have been a consequence of coactivation of the pre- and postsynaptic neurons and also exhibited some specificity. The Hebbian speculation receives support from these and similar data.

That neurons firing together wire together is, however, not the whole story. Whereas neurons become associated when being activated repeatedly at the same time, their antiphasic activation results in weakening of their influence on each other (Tsumoto, 1992). Thus, it appears that it is the *correlation* of neuronal firing of connected cells, or a related measure, which is, so to speak, translated into their connection strength. The correlation c of the firing F of two neurons α and β can be calculated as a function of their probability to fire together in the same small time window, $op[F(\alpha), F(\beta)]$, divided by their individual likelihood to become active, $p[F(\alpha)]$ and $p[F(\beta)]$. For example, the following formula can be used for calculating the correlation:

$$c = \frac{p[F(\alpha), F(\beta)]}{p[F(\alpha)] * p[F(\beta)]}$$

If the strength or weight of a synaptic connection between neurons is determined by the correlation coefficient, this implies (i) that neurons that fire together frequently (high correlation) become more strongly associated, and (ii) that neurons firing independently of each other (low correlation) will be less strongly connected. Connection strength between pyramidal neurons in the cortex is likely strongly influenced by the correlation of their neuronal firing or a closely related measure.

Modification of the strength of synaptic connections can be related to biochemical and even structural changes in the neurons (Kandel, 1991), for example, to the growth and modification of dendritic spines (Braitenberg & Schüz, 1998). It should also be mentioned that coactivated of neurons may lead to formation of new synapses and spines (Engert & Bonhoeffer, 1999). It should therefore be clear that, although the rule proposed by Hebb is important, it is not the only factor determining functional connectivity in the cortex.

2.3 Principles and Implications

The human cerebral cortex is a network of more than 10 billion neurons. Each neuron represents an information processor whose output is a function of the input from the many other neurons with whom it is interwoven. Based on the findings discussed so far, the following principles appear to reflect universal neuroanatomical and neurophysiological properties of the cortex:

- (I) Afferent and efferent projections are ordered.
    - (a) They reach, or insert from, primary areas (Figure 2.3).
    - (b) Sensory and motor projections are organized topographically (Figure 2.4).
- (II) Intracortical connections permit mixing of afferent and efferent information.
    - (a) Adjacent neurons are heavily connected and form *local clusters*.
    - (b) Primary areas tend not to be linked directly, but through *relay areas*.
    - (c) Adjacent areas are connected with high probability (>70%).
    - (d) There is a lower but still good chance for connections between areas farther apart (15–30 percent).
    - (e) Homotopic areas of the two hemispheres tend to be connected.
    - (f) Connections between areas tend to be reciprocal.
- (III) Synaptic connections between neurons are modified depending on their activity.
    - (a) Neurons that fire together strengthen their mutual connections.
    - (b) Neurons that fire independently of each other weaken their connections.

Clearly, this list of structural and functions features of the cortex is not complete and more detail could easily be added. However, these basic facts already make it possible to address more fundamental questions about the function of the cortex.

The cortex is supplied with information ordered according to modality and, within each modality, ordered topographically. The cortex itself now provides multiple indirect links between sensory- and action-related neurons. These links are characterized by great *divergence*: Each neuron reaches thousands of others and each area reaches tens of its sister areas. They are also characterized by great *convergence*; that is, each neuron receives input from multiple other neurons, and each area receives input from several other ones. It has been argued by neuroanatomists (Braitenberg, 1978b; Braitenberg & Schüz, 1998) and neurocomputational modelers (Palm, 1982, 1993a) that this architecture is ideal for *mixing and merging information immanent to sensory- and action-related activity patterns*. Because the mapping between primary areas is indirect, through relay neurons and areas, it is possible to store complex relationships between input and output patterns.

The correlation learning principle (III) implies that frequently cooccurring patterns of activity can be stored by way of strengthening of synaptic links between the participating neurons. Such synaptic strengthening can certainly take place between closely adjacent cortical neurons, but there is reason to assume that such associative learning is also possible between neurons in distant areas. Even coactivation of neurons in different primary cortices causing additional neuronal activity of neurons in their connecting relay areas may be stored by way of increasing synaptic weights between the simultaneously active units. This suggests that webs of neurons can form that are distributed over various cortical areas.

In summary, it appears that the cortex can serve the function of merging multimodal information. This merging of multimodal information is not done by direct links between primary areas, but necessitates intermediate neuronal steps. The intervening neurons between sensory and motor neurons in the cortex allow for complex mappings of information patterns between modalities. Such mappings can be stored by correlation learning.

2.4 Functional Webs in the Cortex

The cortex is a network of neurons characterized by ordered input and output connections in modality-specific areas, by heavy information mixing through short- and long-distance connections, and by correlation learning. Such a device can serve the function of linking neurons responding to specific features of input patterns and neurons controlling aspects of the motor output. Because different primary areas are not linked directly, additional neurons in nonprimary areas are necessary to bridge between the ordered input and output patterns. The cortical connection structure, characterized by a high connection probability between adjacent areas and more selective long-distance links, allows for the formation of functionally coupled but distributed webs of neurons reaching from the primary areas into higher-order cortices. Development of these networks, called *functional webs*, would be driven by sensorimotor or sensory–sensory coactivation, and would be determined by the available cortical projections indirectly connecting the coactivated neurons in primary areas to each other.

2.4.1 Why Numerous Neurons Should Cooperate

It was pointed out by Donald Hebb (1949), and this may be his most important contribution to the understanding of the brain, that synchronously activated neurons should link into cell assemblies, and that cell assemblies underlie all higher cognitive processes. Hebb's proposal diverged radically from earlier neuroscientific approaches to information processing in the brain, because he postulated that higher brain processes are realized as functional units above the level of the neuron. Earlier proposals had stated that either individual neurons (Barlow, 1972) or mass activity and interference patterns in the entire cortex (Lashley, 1950) are the basis of cognition. Hebb's view may appear as a compromise between these views (Milner, 1996).

While Lashley's proposal can be ruled out by considering the specific neuropsychological deficits caused by focal brain lesions (Shallice, 1998), one may ask why large neuron ensembles should become involved in cognitive processing if single neurons are already capable of performing the relevant computations. A tentative answer is that individual neurons are too noisy and unreliable computational devices so that it is advantageous to use sets of neurons working together in functional units to achieve more reliable information processing. If the signal-to-noise ratio of individual neurons is low, one can obtain a better signal by simultaneously averaging a larger number of neurons with similar functional characteristics, so that uncorrelated noise is cancelled (Zohary, 1992). (This does not rule out the possibility that, apart from their shared function, individual neurons in the ensemble can have additional specific functions.) It would therefore make good sense if there were functional units in the cortex that are larger than the neuron but much smaller than the neuronal populations in the macroscopic cortical gyri and sulci.

A further argument in favor of functional webs composed of numerous neurons comes from an estimate of the number of neurons necessary for carrying out the tasks the cortex seems to be primarily engaged in. As mentioned earlier, the cortex includes >10 billion neurons. The number of to-be-stored items can be estimated on the basis of the units that need to be stored. To speak a language well, one needs a vocabulary of fewer than 100,000 words or meaningful language units, called morphemes, and a limited set of rules governing their serial order (Pinker, 1994). Given similar numbers of distinct representations also developed for other cognitive domains, the number of to-be-organized engrams may be in the order of a few 100,000. If this estimate is correct and each engram is represented by one neuron, one million individual neurons might be sufficient for representing the various percepts and motor programs cognitive processes operate on. This raises the question why there are 100,000 to 1,000,000 times as many neurons as would be necessary, as these considerations would suggest. A possible answer is that the cortex includes so many neurons, because individual engrams are realized as populations of neurons of 10^5 to 10^6 neurons.

2.4.2 The Need for Connecting Neurons in Distant Cortical Areas

Functional units in the cortex above the level of single cells are the already mentioned local clusters of neurons beneath approximately 0.1–0.5 mm^2 of cortical surface that, in various sensory areas, respond to similar stimuli

(Hubel, 1995). From the perspective of cognitive science, however, these local neuron clusters *per se* cannot be the substrate of the linkage between different features of an object. The different features of an object may characterize input from different modalities, as, for example, the shape, smell, purr, and smooth fur of a cat. The binding of these features into one coherent representation could be instantiated by pathways linking the sensory information from different modalities to the same "central" neuron(s). These critical cells should then be housed in areas where inputs from many sensory fields converge. It is, however, not necessary to assume a *cardinal cell*, or central convergence area.

As noted, Barlow (1972) has postulated such "grandmother" or "cardinal cells" and Damasio (1989) has proposed that categories of knowledge are processed not by single neurons, but in single "convergence zones" in the cortex. Although these proposals are certainly helpful for understanding aspects of brain function, they do not exhaust the range of possible brain mechanisms. The neuroanatomical connection pattern of the cortex indicates that links between primary cortices are provided through more than one route, involving several nonprimary areas. There is, therefore, no need to assume single specialized areas or neurons for binding of defined entities. Therefore, a distributed web of neurons that processes a certain type of information may include a large number of specialized cells distributed over several cortical areas.

2.5 Defining Functional Webs

A web of neuronal links strongly connecting all neurons involved in the specific processes triggered by an object in the input may become the cortical representation of this object. In this case, binding of object features would be established by mutual links within a *distributed functional web*, that is, between neurons in widespread areas including the primary areas. Each neuron member of the web would therefore contribute to holding the web together thereby playing an essential role in its functioning. The cat concept would be realized as a large set of neurons distributed over a small set of cortical areas. All of these areas would serve as binding sites. A functional web will be assumed to be a set of neurons

(i) That are *strongly connected* to each other
(ii) That are *distributed* over a specific set of cortical areas
(iii) That work together as a *functional unit*
(iv) Whose major parts are *functionally dependent* on each other so that each of them is necessary for the optimal functioning of the web.

The term *functional web* is usually preferred to the term *cell assembly* in the context of this book. The established term *cell assembly* is sometimes avoided because it has been used in so many different ways by different researchers that misunderstandings appear unavoidable (see also Chapter 8). In contrast, the use of the new term *functional web* has now been clarified so that misunderstandings should be unlikely.

To speak about a selection of neurons without implications such as (1) to (4), the expression *neuronal group* is sometimes used. The members of a neuronal group would therefore not need to be connected to each other. Established terms such as *cell assembly*, *neural assembly*, or *neuronal ensemble* are used without the specific implication that (4) is necessarily true. In addition, it is argued that functional webs can exhibit different types of activity states, ignition and reverberation (Section 2.8). This is not a necessary assumption but is, as is argued, supported by empirical data. In Chapter 10, more specific types of functional webs are proposed that may be relevant for grammar processing in the brain.

Whenever one tries to define exactly a given cell assembly or functional web within a simulated or real neuronal network, this turns out to be difficult. In simulations using artificial associative memories (Palm & Sommer, 1995), one immediately encounters the problem of determining which neurons belong to the assembly and which do not. What one finds are neurons that are connected to many of their fellows by maximal synaptic strength. For these, it is out of question that they are assembly members. However, there are also others whose connections to the rest of the assembly is slightly weaker and whose inclusion in the assembly is therefore uncertain. Correspondingly, in network simulations and also in neurophysiological observations, some neurons almost always become active together, but others may only be recruited together with them with lower probability, depending, for example, on the neuronal sets activated in the past (Milner, 1957). In other words, there are neurons exhibiting high correlation among each other and others whose correlation with these “core neurons” is smaller. To decide whether a neuron is part of an assembly or web, critical values for synaptic connectedness or correlation coefficients must be introduced. Definition of such critical values is always arbitrary. For some purposes, it helps to distinguish the *kernel* or *core* of an assembly from its *halo* or *periphery* (Braitenberg, 1978a), but in network simulations, necessarily arbitrary boundaries must also be introduced for defining these assembly parts.

Cell assemblies and functional webs are necessarily fuzzy concepts. This does not constitute a principled problem. It is essential to see that fuzziness is immanent to the assembly concept and that this is problematic in exactly the same way as it is difficult to determine the boundaries of the sun or the milky way. The fuzziness of the boundaries of the respective concepts should not obscure the fact that there are stars, galaxies, and perhaps functional webs in the brain that may represent these concepts and words.

Perhaps most importantly in the context of a cell assembly approach to language, the meaning of words in ordinary language use also is characterized by the lack of sharp boundaries (see Chapter 5; Wittgenstein, 1953). If our conceptual system is fuzzy, it is probably best to model it using fuzzy mechanisms and to speak about it using words with fuzzy meaning. The anatomical fuzziness of the boundaries of the individual functional web should not obscure the fact that, functionally, each web is conceptualized as a coherent unit, a *discrete entity*.

2.6 Evidence for Functional Webs

Which critical predictions are implied by the idea of distributed functional webs? If the cat concept is neuronally represented as a distributed web of neurons that form a functional unit, this has two important implications:

(1) A significant portion of the web's neurons are active whenever the cat concept is being processed.
(2) The function of the web depends on the intactness of its member neurons.

If neurons in the functional web are strongly linked, they should show similar response properties in neurophysiological experiments. If the neurons of the functional web are necessary for the optimal processing of the represented entity, lesion of a significant portion of the network neurons must impair the processing of this entity. This should be largely independent of where in the network the lesion occurs. Therefore, if the functional web is distributed over distant cortical areas, for instance, certain frontal and temporal areas, neurons in both areas should (i) share specific response features and (ii) show these response features only if the respective other area is intact.

These predictions have been examined in macaque monkeys using a memory paradigm in which the animal must keep in mind the shape or color of a stimulus and perform a concordant matching response after a delay of several seconds (delayed matching to sample task). Throughout the memory period, in which the animal must keep in mind, for example, that the stimulus shown was red, neurons fire at an enhanced level. Their firing is specific in the sense that they do not respond, or respond less, when a stimulus of a different color is shown. Neurons with this stimulus-specific response pattern were found in the prefrontal cortex (Fuster & Alexander, 1971) and in the inferior temporal lobe (Fuster & Jervey, 1982). Memory neurons in different cortical areas showed similar dynamics. After their activation, there was a near-exponential decrease of neuronal activity in many of them. Thus, prediction (i) is met: Neurons in different cortical lobes showed similar response properties.

Another facet of prediction (i) is that the participating neurons, some of which may play a genuine role in the control of action (frontal lobe) and perception (temporal lobe), share their functional properties after the distributed network has been set up. The neurons in temporal lobes may therefore show functional characteristics of neurons involved in perception and, as a result of their strong links to action-related neurons in the frontal lobe, they may also share their functional characteristic of firing in the context of certain actions. They would, therefore, show multimodal response characteristics. The same line of thought would apply for neurons in the frontal lobe with a genuine role in action control and, in addition, a strong linkage to temporal neurons supplying them with sensory information. Evidence on cells with multimodal response properties (Fuster, Bodner, & Kroger, 2000; Rizzolatti, Luppino, & Matelli, 1998; Zhou & Fuster, 2000) strongly support such multimodal response characteristics of cortical cells – in particular, neurons in the premotor and prefrontal cortex. The idea that the cortex is an information mixing device and the concept of distributed functional webs allowing single neurons to represent and process merged information from various motor and sensory modalities receive further support from these data. Mirror neurons (Rizzolatti & Arbib, 1998; Rizzolatti et al., 1996) that have a role in controling an action and also respond to the perception of the execution of this action are important constituents of the postulated functional webs.

Prediction (ii) was that lesions in frontal areas on the one hand, and temporal or other posterior cortices on the other hand, should impair both the perceptual and action-related aspects of higher cognitive processes. This should not apply for very elementary perceptual processes, such as the perception of basic stimulus features or elementary movements, but should apply for cognitive processes such as those involved in memory and language. In memory experiments (delayed matching to sample tasks) performed with macaque monkeys, neurons in frontal and inferior temporal areas showing similar stimulus- and action-related activity were investigated. Temporary lesion by cooling of the neurons in frontal areas led to loss of stimulus specificity of the neurons in the temporal areas, and cooling in the frontal areas also impaired the stimulus specificity of the neurons in the temporal cortex. Thus, temporary lesion of stimulus-specific or, more precisely, engram-related neurons in one area led to functional impairment of the neurons in the respective other area (Fuster, 1997). Together, these data provide evidence that neurons in both temporal and frontal areas (a) showed the same specific response features and (b) showed these response features if and only if the respective other area was intact (Fuster, 1997).

These results obtained in memory experiments with macaque monkeys are reminiscent of well-known facts from more than 100 years of neurological investigation into acquired language disorders, aphasias (Broca, 1861; Lichtheim, 1885; Wernicke, 1874). These classical studies of aphasia showed that prefrontal and temporal areas are most crucial for language processing. They also showed that lesions in either area can lead to aphasia, which in the majority of cases include deficits in both language production (Lichtheim, 1885) and perception (De Renzi & Vignolo, 1962). Concordant with recent animal studies investigating the consequences of local cooling of prefrontal and temporal areas, this suggests mutual functional dependence between frontal and temporal areas (Pulvermüller & Preissl, 1991).

2.7 A View of Cortical Function

The reviewed facts from basic neuroscience make it likely that the cortex includes distributed neuron ensembles involving neurons in cortical areas in different lobes that show similar response properties and whose intactness is necessary for defined cognitive operations. This further advocates the following view of cerebral cortical function. The cortex is an associative memory allowing for merging information from different modalities. Merging of information is driven by correlation of spatiotemporal patterns of neuronal activity carrying information about sensory perceptions and actions. These correlated activity patterns occur in action-related frontal and sensory-related posterior areas of the cortex. The participating neurons are being bound into strongly connected webs of neurons, functional units that represent cognitive entities with sensory and action aspects (words, concepts engrams in general).

This view on cortical function leads to further questions, as follows:

(1) *What* are the functional dynamics of these distributed neuronal representations?
(2) *Where* exactly are they localized, or, formulating the question in a slightly more adequate manner: Over which cortical areas are these functional webs distributed, i.e., what is their cortical topography?
(3) *How* can the internal wiring of the functional webs be specified?

(4) Is it sufficient to assume that the formation of these networks is driven by associative learning principles, or do genetically determined factors play a role as well?

The "What" question (1) about functional dynamics is answered tentatively in the last section of this chapter and in Chapters 10 and 12. The "Where" question (2) is addressed in detail in Chapters 4 and 6. The "How" question (3) receives a tentative answer in Chapter 8 on serial order, and the last question (4) about the possible role of genetically determined information is likely impossible to answer, but an attempt is made in Chapters 4 and 14.

2.8 Temporal Dynamics in Functional Webs: Ignition and Reverberation

On theoretical grounds, one can predict that a strongly connected neuron population produces overshooting activity if strong stimulation reaches it from the outside. This would mean that, although only a fraction of the neuron members of a web are stimulated, all members of the web, or at least most of them, would nevertheless become active, and activity might even spread to other networks as well. Restricting considerations here to the activity dynamics within the functional web, stimulation of a fraction of its neurons can lead to a full activation of the entire population. This process has been called *ignition* (Braitenberg, 1978a). Neural network models using associative memory models leave no doubt that such a process takes place in a sufficiently strongly connected and sufficiently strongly stimulated network of neurons (Palm, 1981, 1982).

If activation of a substantial fraction of the neurons of a functional web leads to its ignition, one may consider this an undesired consequence of the architecture of network memories. However, the assembly internal spread of activity may as well be considered as positive because the activation of the web, so to speak, completes itself as a result of the strong web-internal links. If the web of neurons is considered a memory representation of an object and each neuron to represent one particular feature of this object memory, the full ignition would be the neuronal correlate of the activation of the stored object representation. Such full activation of the object memory could occur if only a fraction of the features of the object are present in the actual input. This could be a psychologically important process, an organic correlate of the completion of a gestalt in the perception process. Gestalt completion can be modeled in associative memories (Willshaw, Buneman, & Longuet-Higgins, 1969), and there is reason to believe that a similar process is relevant in the real brain.

The dynamics of a functional web would therefore be characterized by full ignition of the web after sufficiently strong stimulation by sensory input, or, as an alternative, after stimulation by cortical neurons outside the functional web. This latter cortex-internal activation of a web can be considered the organic basis of being reminded of an object even though it is absent in the environment.

The ignition process likely takes place within a short period of time after the functional web has been sufficiently stimulated. An exact definition of the interval is probably not sufficiently motivated by empirical data, but an educated guess might be that the ignition occurs within 200 ms after sufficient information is present in the input. The issue of temporal dynamics of memory representations is further discussed in the context of neurophysiological activity accompanying word processing (see Chapter 4).

When an ignition has taken place, refractory periods of neurons and other fatigue effects may again reduce the activity level of the functional unit. It is also possible that a regulation process designed to keep the cortical level of activity within certain bounds (see Chapter 5) becomes active and reduces activity by global inhibition or removal of background activity (disfacilitation). In this case, however, activity will not necessarily completely die out in the ensemble of neurons that just ignited. The strong within-assembly connections may still allow activity to be retained in the neuron set. This is a putative neurobiological basis of *short-term* or *active memory* (Fuster, 1997, 1998a, 1998b, 2000). The distributed cortical functional web itself would therefore be the organic side of a long-term (or *passive*) memory trace, and the sustained activity of the same web would realize the short-term (or *active*) memory.

Researchers who have found memory-specific neurons in the neocortex have investigated the dynamics of these cells extensively while the experimental animal must keep a certain piece of information in mind. The animal saw a stimulus of a certain color and had to push a button of the same color after a memory period of 20 seconds. Throughout the entire memory period, neurons were found that stayed at an enhanced activity level. Some of these neurons were stimulus specific, therefore responding strongly only if the to be remembered stimulus exhibited a particular feature. The neuron did, for example, respond strongly to a yellow stimulus, but did not show any enhanced activity, or at least responded significantly less, when the stimulus had a different color, for instance, green or blue (Fig. 2.6). The explanation is that ensembles of neurons were activated by the input and stayed active during the memory period. The activity dynamics of the individual memory cell would be assumed to reflect the activity dynamics of the memory network. The enhanced activity level of the neuron and web would be the

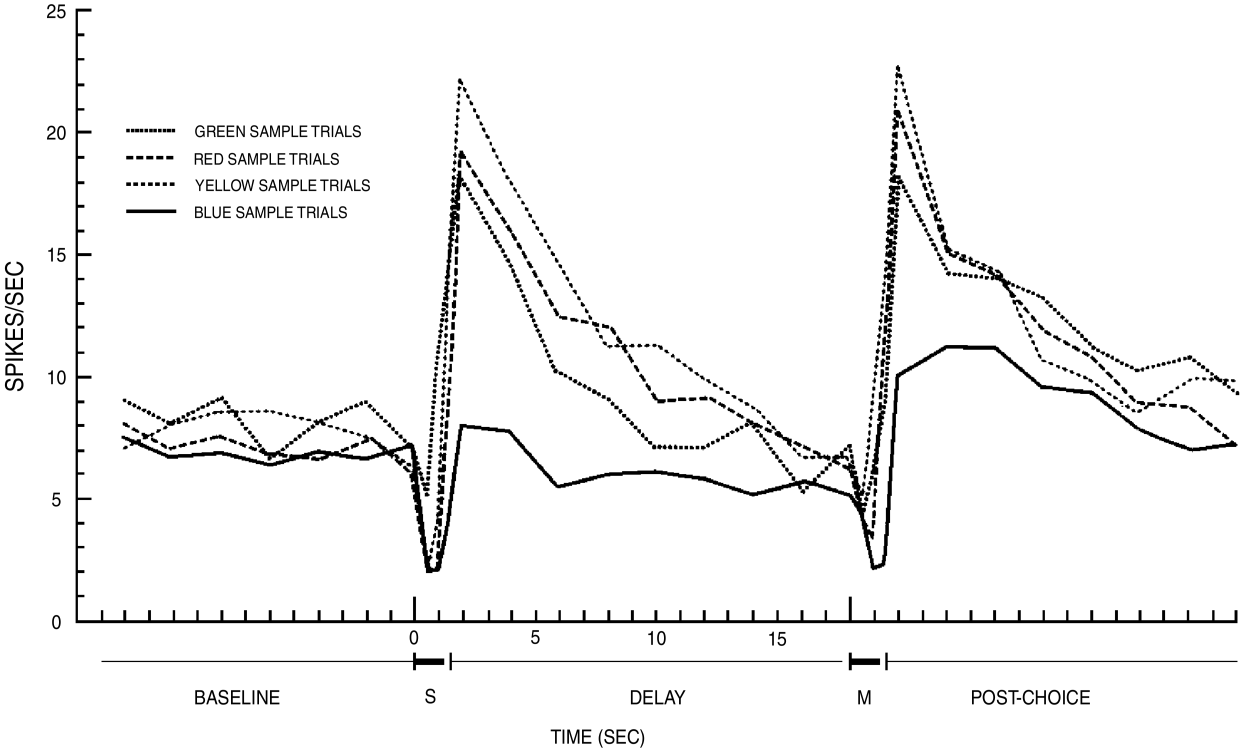

Figure 2.6. Result of an experiment in which a monkey had to keep in mind if a green, red, yellow, or blue sample stimulus was shown during the stimulation interval S. During the memory period ("delay"), the neuron from which recordings were taken responded maximally after a yellow sample stimulus. High activity of the neuron appears to be an indicator of a specific engram in memory. During the delay, the neuron loses activity almost exponentially. Reprinted with permission from Fuster, J. M. (1995). *Memory in the cerebral cortex. An empirical approach to neural networks in the human and nonhuman primate*. Cambridge, MA: MIT Press.

result of the reverberation of activity within the functional web. The stimulus specificity of the neurons suggests memory networks specific to objects or object features.

This research revealed sometimes complex activity dynamics. As mentioned earlier, neurons usually did not stay at a constant enhanced level of activity; instead, gradual activity changes were found. Many neurons showed strong activation after stimulus presentation followed by an exponential decline of activity with time throughout the memory period. Figure 2.6 illustrates this. The activity during the memory period – in action potentials per second – is displayed as a function of time when stimuli of different color had to be memorized. These data can be interpreted as support for the idea that after the ignition of a cortical neuron ensemble corresponding to the stimulus, this network stays at an enhanced activity level for several tens of seconds. It has been proposed that activation of a distributed cortical neuron ensemble is reflected by coherent oscillatory activity of the participating neurons. This claim receives support from both neurophysiological studies in animals (Kreiter, 2002; Singer et al., 1997) and EEG and MEG studies in humans (Pulvermüller, Birbaumer, Lutzenberger, & Mohr, 1997; Tallon-Baudry & Bertrand, 1999).

In summary, neurophysiological data and neurocomputational investigations suggest complex activity dynamics of functional webs. An initial brief ignition of the assembly appears to be followed by a reverberatory state in which the assembly retains activity, although the level of activity may fall off exponentially with time. This memory interval of reverberatory activity can last for tens of seconds.

CHAPTER THREE

From Classic Aphasia Research to Modern Neuroimaging

In this chapter, an introduction to facts known from neurological, neuropsychological, and neuroimaging research on language is given. How the obtained results can be explained and how they fit together is discussed.

3.1 Aphasiology

The scientific study of language proliferated in the second half of the nineteenth century. Apart from linguists and psychologists, neurologists focused on language in the context of the then new findings about "affections of speech from disease of the brain" (Jackson, 1878). In adults who had been fully able to speak and understand their native language, a stroke, tumor, trauma, or encephalitis was sometimes found to severely and specifically reduce their language abilities. Such language disturbances were called *aphasias*. There was, and still is some discussion as to whether there are subtypes of aphasia, and a good deal of the research on aphasia was dedicated to developing new classification schemes and arguing why one scheme should be a better reflection of what appeared to be the truth than another. The research effort resulted in numerous classification schemes (Caplan, 1987) and also in what appeared to be an extreme position expressed by the claim that there is only one type of aphasia (Marie, 1906).

This last view can be based on the common features of all – or at least the large majority of – aphasias. All aphasics have difficulty speaking, although their ability to move their articulators – lips, tongue, pharynx, larynx, and other muscles in the mouth-nose region – may be well preserved. When confronted with spoken and written language, all aphasics exhibit some difficulty. An easy way of finding out whether an individual suffering from brain disease has an aphasic disturbance is to present him or her with a few simple commands and assess whether the patient is still able to carry them out. A test called the *Token Test* (De Renzi & Vignolo, 1962) includes a list of commands to manipulate colored tokens of different shape and size. If a patient has an aphasia, he or she usually shows difficulty understanding and following commands such as, "Please touch the little yellow circle," or, "Please take all the squares except for the red one." In addition to their comprehension problems, all aphasics have difficulty producing spoken and written language output. It has been reported that there are exceptional patients who would show a deficit only in oral language production, or a deficit only in comprehending certain aspects of spoken language (Kolk, van Grunsven, & Keyser, 1985). Apart from these few exceptions, however, aphasias appear to be *multimodal*; that is, they include deficits in producing and comprehending spoken language and additional deficits in reading and writing as well. The multimodal character of aphasia can be used to argue in favor of the view that there is only one type of aphasia, aspects of which happen to be more or less pronounced in different cases.

However, if the focus is on the diversity of the actual problems patients with acquired language deficits exhibit, it makes sense to distinguish different syndromes of aphasia. The most obvious differentiation is between aphasics whose primary difficulty is producing language and aphasics who show a most massive comprehension deficit. Here, a distinction between *motor aphasia* and *sensory aphasia* has been proposed. Today, the terms *Broca aphasia* and *Wernicke aphasia* – referring to these variants of the disease, respectively, and to the two neurologists who first described them – are established. Apart from these two aphasic syndromes, a severe form of aphasia, *global* or *total aphasia*, and a slight disturbance characterized mainly by word-finding problems, *amnesic aphasia* or *anomia*, are sometimes distinguished. More fine-grained syndromes of aphasia have been proposed on the basis of theoretical approaches of language processing in the brain.

In the second half of the nineteenth century, neurologists proposed *connectionist theories* to summarize and model the effect of brain lesions on cognitive functions. The main idea underlying these models was that local processing systems, or *centers*, specialize in particular cognitive processes. These were believed to be autonomous processors realized in large portions of gray matter, such as the superior temporal gyrus or the inferior frontal gyrus. The centers were thought to be linked through *pathways* allowing information exchange between them. The pathways in the models had their analog in fiber bundles in the white matter of the brain.

The most famous classical neurological model of language goes back to Broca (1861) and Wernicke (1874) and has been proposed by Lichtheim (1885). This model is consistent with basic features of aphasia. The main proposal of Lichtheim's model was that there are two centers for language

Table 3.1. Nomenclature of some areas relevant for language processing. Names of areas are indicated and related to the number code proposed by Brodmann. (See Fig. 2.3 for the corresponding loci in Brodmann's area map.)

Name	Brodmann number	Latin name	Location
Broca's area	44	Gyrus opercularis	Inferior frontal lobe
	45	Gyrus triangularis	Inferior frontal lobe
Wernicke's area	22	Gyrus temporalis superior	Superior temporal lobe
	39	Gyrus supramarginalis	Inferior parietal lobe
	40	Gyrus angularis	Temporo–parieto–occipital junction

processing: one involved in speech production and the other in language comprehension. The existence of two separate anatomical centers relevant for language, the areas of Broca and Wernicke, has meanwhile been confirmed by many clinical observations. These areas are located in the left hemisphere in most individuals, Broca's area in the inferior frontal gyrus, Brodmann areas 44 and 45; and Wernicke's area in the superior temporal lobe, Brodmann area 22. Both are located close to the Sylvian fissure, that is, in the *perisylvian region* (Table 3.1).

Figure 3.1 presents Lichtheim's model of language representation and processing in the brain. He distinguished three centers, two of which are devoted exclusively to the processing of aspects of language. The first center, the motor language center M, was supposed to store what Lichtheim

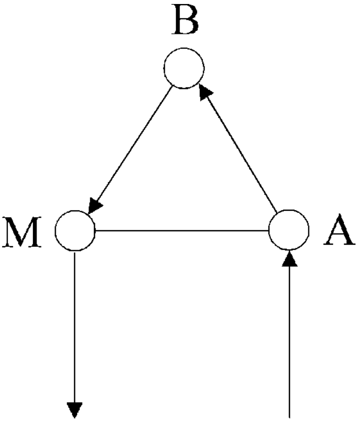

Figure 3.1. Lichtheim's model of language in the brain. Language was proposed to be realized in three centers involved specifically in representing and processing sound patterns of words (*Wortklangbilder*; acoustic language center A), articulatory patterns of words (*Wortbewegungsbilder*; motor language center M), and concepts (*Begriffe*; concept center B). Adopted from Lichtheim, L. (1885). Über Aphasie. *Deutsches Archiv für Klinische Medicin, 36*, 204–68.

called *Wortbewegungsbilder*, or representations of the articulatory movements performed to pronounce a word. The second center, the auditory language center A, was thought to store what Lichtheim called *Wortklangbilder*, or images of the sound sequences of words. The third center, which was hypothesized to play a major role in other mental activities as well, was called the *Begriffszentrum*, the center storing concepts, B. Any two of these three centers were proposed to be connected by pathways allowing for converting a sound sequence directly into a sequence of muscle movements or a concept into a word and back. The genuine language centers A and M correspond to the areas of Broca and Wernicke.

Language processes were modeled by activity flowing through the connectionist architecture. For example, word comprehension would be modeled as the activation of A by acoustic input and the subsequent activation of B through its connection to A. The cortical process of language production, however, was mimicked by activity in B, exciting, through the B–M pathway, the motor center M, and finally causing a motor output. The process of thoughtless repetition of a word was believed to be possible without activating B, based on the direct connection between the acoustic and the motor centers. Aphasias were now modeled as a lesion to one or more centers or pathways. Broca and Wernicke aphasias, for example, were realized in the model as lesions of the respective centers M and A.

The nice thing about Lichtheim's scheme was that it suggested additional syndromes of aphasia that could be confirmed empirically. Examples are *conduction aphasia*, modeled as lesion of the connection from A to M, a syndrome primarily characterized by difficulty in repeating word sequences, and a mirror-image syndrome called *mixed transcortical aphasia*, in which repetition is still relatively intact whereas all other language abilities including the ability to understand what is being repeated are severely impaired.

Although the existence of two centers whose lesion in many cases leads to severe language disturbance could be confirmed, it is less clear that these centers have the specific function Lichtheim and Wernicke attributed to them. As Lichtheim recognized in 1885, it is incorrect to assume that what has been labeled the "center for language comprehension" is relevant only for this single language function. As emphasized earlier, patients with Wernicke aphasia have difficulty speaking, even if their lesion is restricted to the superior temporal lobe (Pulvermüller, Mohr, Sedat, Hadler, & Rayman, 1996). They use words in inappropriate contexts; produce words incorrectly, with incorrect language sounds in them or language sounds omitted; or even produce an incomprehensive jargon of mixed-up language sounds. These deficits are typical for subtypes of sensory or Wernicke aphasia, and cannot be easily explained by assuming a selective lesion to a center devoted to language comprehension. Furthermore, the neurologically caused speech errors frequently made by Wernicke aphasics differ qualitatively from speech errors made by Broca aphasics, who tend to speak very slowly because they cannot produce words they intend to use, rather than replacing them with wrong ones. A lesion in an "auditory language center" cannot account for such differential output. A similar argument can be made for the comprehension side, because in Broca aphasia, specific comprehension deficits are particularly apparent when patients are confronted with certain sentence types, including, for example, passive sentences (Caramazza & Zurif, 1976). Thus, it would be incorrect to postulate a cortical center specifically for language production and a second independent center processing auditory language input exclusively. The two areas most crucial for language processing in the cortex, the inferior frontal area of Broca and the superior temporal area of Wernicke, appear to be functionally interdependent (see Chapter 2; Excursus E1).

Why are the language areas located where they actually are? It might be tempting to propose an additional principle, adding to the ones proposed earlier in Chapter 2; a principle saying that the human cortex is constructed such that language must be processed in these very areas. As an alternative, however, it is possible to explain the location of the language centers by principles already known from neuroanatomy and neurophysiology.

It is clear that production of a language element, for example, a syllable or word, corresponds to activation of neurons controlling movements of the articulators. It is likewise evident that such articulations must always activate neurons in the acoustical cortical system, unless the auditory system is damaged or otherwise dysfunctional. Thus, when uttering a language element, there must be correlated neuronal activity in the perisylvian motor cortex and auditory cortex in the superior temporal lobe. Because between-area connections preferably connect adjacent areas, this neuronal activity can spread to adjacent fields in the superior temporal and inferior frontal lobes. These areas are connected by long-distance connections. Thus, a sufficiently strong correlated activity pattern in primary motor and auditory areas gives rise to the activation of a specific neuron population spread out over these inferior frontal and superior temporal areas, including neurons in the language areas of Broca and Wernicke. The principles outlined in Chapter 2 are sufficient for explaining why the major language areas are where they are.

The outlined view of the development of sensory-motor links in language acquisition has been established for some time (Braitenberg, 1980; Fry, 1966). It is sometimes called the *babbling hypothesis*, because during the infant's repeated articulation of syllables within the first year of life, this type of associative learning may occur for the first time. Strong associative links between the two principal language areas predict that these areas are not functionally independent, but that they become active together and cooperate when they generate, or respond to, language sounds, words, or sentences. This cooperation postulate contrasts with the view immanent to the Lichtheim scheme that the frontal and temporal language areas are autonomous centers. Strong associative links between the areas of Broca and Wernicke may be the basis of the multimodal character of most aphasias, a feature Lichtheim's scheme cannot account for. This issue was touched upon earlier and will be discussed further in detail in Excursus 1.

One may object to this view by mentioning that in users of sign language, large lesions in the perisylvian areas also lead to aphasia (Poizner, Bellugi, & Klima, 1990). Should not the frontal language center for sign languages be located superior or dorsal to that of spoken language because signing is done primarily by moving one's hands rather than the face and articulators? Recall that the somatotopic organization of the primary motor cortex implies a more dorsal position at least for the neurons directly controlling arm movements (see Fig. 4.4). However, because many of the lesions seen in deaf signers with aphasia were rather large and affected both the perisylvian areas and the more dorsal areas controlling hand movements, they cannot help decide the issue. There are very few exceptions – for example, the case of a patient called Karen L., who suffered from a relatively small lesion dorsal to Broca's area but which extended into the supramarginal and angular gyri (Poizner et al., 1990). Poizner and colleagues emphasize that a lesion of this kind does not usually lead to persistent aphasia including comprehension deficits, and attribute this patient's aphasia to the posterior areas affected. However, it is also possible to relate the deficit to the anterior part of the lesion dorsal to Broca's region. The data on aphasia in signers appear to be consistent with an explanation of the cortical locus of the areas most relevant for language based on correlated neuronal activity and the available neuroanatomical links.

Unfortunately, neuroimaging data addressing the issue of the cortical areas involved in sign language processing are controversial at present. Some colleagues report massive differences in the cortical organization of spoken and signed language, for example in the pattern of cortical laterality (Bavelier, Corina, & Neville, 1998; Neville et al., 1998; Newman et al., 2002). In contrast, others claim that the cortical locus of spoken language and that of sign language are similar (Petitto et al., 2000). The interpretation of brain correlates of spoken and sign languages is complicated by cortical reorganization (Buonomano & Merzenich, 1998). If arms and hands rather than the articulators are excessively used for communicating, general cortical reorganization processes may be triggered so that some of the articulators' cortical space is taken over by the extremities. This may explain why similar areas are activated in the frontal lobe when signed and spoken languages are processed (Petitto et al., 2000). Cortical reorganization may be an important factor working against clearly separate cortical loci devoted to the processing of spoken and signed languages.

Whereas the language areas of Broca and Wernicke were for some time believed to be the only areas crucial for language, neuropsychological work carried out in the last quarter of the twentieth century proved that also other areas are necessary for unimpaired language processing. In particular, lesions in the frontal and temporal lobes, some of which spared the perisylvian language areas, led to difficulty producing or understanding words. Many of these aphasic deficits were most pronounced for words from particular categories – nouns, verbs, or more fine-grained semantic subcategories of words and concepts (Damasio & Tranel, 1993; Humphreys & Forde, 2001; Warrington & McCarthy, 1983; Warrington & Shallice, 1984). For example, frontal lesions appear to limit the ability to process verbs (Bak et al., 2001; Daniele et al., 1994; Neininger & Pulvermüller, 2001), whereas inferior temporal lesions were found to most severely impair the processing of nouns from certain categories (Bird et al., 2000; Damasio et al., 1996; Warrington & McCarthy, 1987). Clearly, these language-related deficits remain unexplained by the Lichtheim scheme.

3.2 Laterality of Language

Laterality of language is a well-known fact since the first scientific investigation of language loss due to stroke (Broca, 1861), but the causes of this laterality have not yet been revealed. The postulate that one hemisphere is dominant for language is based primarily on lesion studies. Only lesions in the left hemisphere cause aphasias in most individuals. It was pointed out by English neurologist Hughlings Jackson (1878) that if a lesion of a part of the brain impairs specific functions, one can by no means conclude that these functions are localized in the respective brain part. The lesioned area could have a more general function, as the brain stem has in regulating arousal, which is necessary for, but not specific to, a specific higher brain function such as language. In this case, one would perhaps not want to localize language in the brain part in question, although language impairments result from its lesion. Likewise, if lesions of a brain part lead to a clinically apparent deficit regarding a given function, it is always possible that additional areas are also relevant for this function, but that their lesion does not result in clinically apparent dysfunction. Such deficits may be absent, for example, because the clinical tests applied are not sensitive enough to reveal a fine-grained dysfunction (Neininger & Pulvermüller, 2001), or because other areas had meanwhile taken over the damaged area's function (Dobel et al., 2001; Price et al., 2001; Weiller et al., 1995). Therefore, Jackson's early claims are correct: Loss of function F after lesion in area A does not prove that F is exclusively housed in A, and the absence of an F-deficit as measured by clinical tests, after lesion in A does not prove that A has no role in the respective function. Therefore, lesion data proving language laterality do not argue against the existence of additional sites in the nondominant hemisphere that are also relevant for language processing.

Whereas lesions in certain left-hemispheric areas cause severe language impairments and aphasias, right-hemispheric lesions usually lead to more subtle language-related difficulties affecting prosodic and pragmatic processing (Joanette, Goulet, & Hannequin, 1990). In this sense, left-hemispheric language dominance is almost always present in persons who are right handed and also in most left-handed individuals (~80 percent) (Bryden, Hecaen, & DeAgostini, 1983; Goodglass & Quadfasel, 1954; Hecaen, De Agostini, & Monzon-Montes, 1981). The remaining individuals are right dominant, with a few showing no language dominance at all. Therefore, it is obvious that in the large majority of individuals, language is lateralized to the left hemisphere.

Language laterality was also reflected in brain physiology revealed by modern neuroimaging techniques. Stronger brain responses in the left hemisphere in comparison with the right hemisphere were seen across various tasks using visual or auditory stimuli (Petersen & Fiez, 1993). Because lateralized activity was elicited already by single language sounds and syllables (Näätänen et al., 1997; Shtyrov et al., 2000; Zatorre et al., 1992), one may conclude that phonological processes, or more precisely, acoustic processes relevant for the distinction of language sounds are crucial for language laterality. In many of the neuroimaging studies mentioned, in particular in studies using magnetoencephalography (MEG), multichannel electroencephalography (EEG), or functional magnetic resonance imaging (fMRI), language laterality was gradual, in the sense that there were activity signs in both hemispheres and the dominant left hemisphere was more active than the right (Pulvermüller, 1999b).

Neuropsychological and neurophysiological studies indicate that laterality of language is present early in life. Infants suffering from brain lesions are more likely to develop a temporary language disturbance after left-hemispheric lesion than after a lesion to the right hemisphere (Woods, 1983). The great plasticity of the neural substrate allows for recovery in most cases of early neurological language impairment. EEG recordings in young infants demonstrated a physiological correlate of language laterality early in life, even in the first year (Dehaene-Lambertz & Dehaene, 1994; Molfese, 1972).

In which way does the laterality of language functions relate to structural asymmetries? Numerous putative anatomical correlates of language laterality have been reported, even in craniofacial asymmetries during early ontogenetic stages (Previc, 1991). Neuroanatomical Language laterality was proposed to be reflected neuroanatomically in the size of language-relevant areas (Geschwind & Levitsky, 1968; Steinmetz et al., 1991), and the size (Hayes & Lewis, 1993), ordering (Seldon, 1985), local within-area connections (Galuske et al., 2000), and dendritic arborization patterns (Jacobs et al., 1993; Scheibel et al., 1985) of cortical pyramidal neurons. These anatomical differences may have a causal role in determining which hemisphere becomes more important for processing spoken language, although the causal chains are, as mentioned, not yet understood. However, one may well argue that some of the structural asymmetries are a consequence of functional differences – for example, of more strongly correlated neuronal activity in the respective areas (see Chapter 2).

Considering the documented anatomical and functional asymmetries, it becomes important to illuminate the possible causes of left hemispheric laterality of language, and to think about the actual causal chain. According to one view, specific neuroanatomical differences between the hemispheres cause laterality of neurophysiological processes important for distinguishing language sounds, or *phonemes*. Based on an extensive review of the neuroanatomical literature, Miller (1996) found that the ratio of white to gray matter volume yields a smaller value for the left hemisphere in comparison to the right hemisphere. The structures most crucial for language, the frontal and temporal lobes, exhibit a smaller volume of white matter in the left hemisphere. Thus, a smaller white matter volume appears to be related to language dominance. The white matter is made up primarily of axons and their glia sheaths, the long-distance cables connecting cortical neurons. A smaller white matter volume may indicate that average cortical connections are thinner, and this implies that these connections conduct action potentials more slowly (Lee et al., 1986). This line of thought leads Miller to propose that the left hemisphere includes more slowly conducting fibers. In local cortical circuits, slow fibers may be advantageous for measuring exact temporal delays, and measuring exact temporal delays in the order of a few tens of a millisecond is necessary for making phonemic distinctions such as between a [t] and a [d]. According to this view, language laterality is phonologically related and a direct consequence of neuroanatomical properties of the human forebrain. However, this theory is, as all other attempts at further explanation of language laterality, in need of more empirical support.

In summary, laterality of spoken language is supported by neuropsychological lesion studies as well as neuroanatomical and neuroimaging results. Although it is a well-established fact, its explanation based on neuroscientific principles is tentative. Because of the obvious need for further explaining this fact on the basis of more fundamental neuroscientific knowledge, it may appear safer to postulate an additional principle underlying cortical language function, which would need to be added to the neuroscientific principles outlined in Chapter 2.

(IV) Language processes are lateralized to the dominant hemisphere.

Such a principle must be postulated until the causes of the laterality of language are better understood in terms of more elementary neuroscientific principles, such as the ones expressed by Principles (I)–(III).

As pointed out, laterality of language does not imply that language mechanisms are housed exclusively in the dominant hemisphere. The possibility exists that the nondominant hemisphere also contributes to language processes and, furthermore, that it contributes specifically to central aspects of linguistic processing. The imaging studies on language laterality, such as those mentioned previously, show that both hemispheres become active during language tasks, but they do not prove that the contribution of the nondominant hemisphere to language processing is crucial. However, neuropsychological work made it possible to show that the right hemisphere alone is capable of word processing. Studies in patients whose cortical hemispheres were disconnected (*split-brain patients*) or in whom the dominant hemisphere was removed (*hemispherectomy patients*) to cure an otherwise intractable epilepsy showed word-processing abilities with their isolated right nondominant hemisphere. The nondominant hemisphere could, for example, reliably distinguish words from meaningless material (Zaidel, 1976, 1985). This has been shown in so-called lexical decision tasks, in which subjects must decide whether a written letter string is a word by pressing a button. If the visual stimulus is presented to the left of fixation, the visual information is being transferred only to the right hemisphere (because the fibers in the visual pathway cross). Presentation to the right of fixation stimulates the left hemisphere. Split-brain and hemispherectomy patients were able to perform well on lexical decision tasks when words were presented in their left visual field. Thus, the right hemisphere (which was nondominant for language) must have processed them. This proves that, apart from the contribution of the nondominant hemisphere to pragmatic aspects of language (Zaidel et al., 2000), its neuronal circuits are capable of word processing as well.

Thus, the imaging and neuropsychological data mentioned show that the nondominant cortical hemisphere is activated during language tasks and is capable of some language processing on its own. However, one may still hold the view that the language mechanisms in the nondominant hemisphere do not become relevant unless there is dysfunction in the main language machinery of the dominant half of the cortex. To prove entirely that nondominant right-hemispheric processes are necessary for unimpaired language processing, one must show (i) that additional involvement of the right nondominant hemisphere improves language processing compared to left-dominant hemispheric processing alone, and (ii) that lesions in the right hemisphere can reduce language processing abilities.

Can the nondominant hemisphere actually assist the dominant hemisphere in word processing? This question was addressed by lexical decision experiments in which stimulus words and pronounceable pseudo-words (for example, “moon” vs. “noom”) were presented to the left or right of fixation of healthy subjects so that information was being transmitted to the contralateral cortical hemisphere. As an additional condition, identical words, that is, two copies of the same word, were shown simultaneously to the left and right of fixation, so that both hemispheres received the information simultaneously. The results showed clearly that bilateral stimulation led to faster and more reliable processing of words compared to stimulation of the dominant left hemisphere alone. This *bilateral advantage* could not be observed, or was substantially reduced, for meaningless letter strings or pseudo-words, and it was also absent in split-brain patients (Mohr, Pulvermüller, Mittelstädt, & Rayman, 1996; Mohr, Pulvermüller, Rayman, & Zaidel, 1994; Mohr, Pulvermüller, & Zaidel, 1994; Zaidel & Rayman, 1994). These data are consistent with the view that additional information processing in the nondominant hemisphere can help the dominant hemisphere process words. The right hemisphere seems to play a role in optimizing word processing (Pulvermüller & Mohr, 1996; Hasbrooke & Chiarello, 1998).

Can lesions in the right nondominant hemisphere lead to linguistic impairments? This question was addressed by confronting patients with words from different categories in a lexical decision task. The result was that patients with lesions in different parts of the right hemisphere showed selective degradation of words from specific categories, such as words related to actions or words related to visual perceptions. These category-specific deficits revealed by the lexical decision task show that an intact right nondominant hemisphere is necessary for unimpaired processing of words from particular categories (Neininger & Pulvermüller, 2001).

In summary, although language laterality is a well-established phenomenon, the nondominant hemisphere contributes to, is sufficient for, and is also necessary for the optimal processing of language.

3.3 Neuroimaging of Language

The knowledge of the cortical basis of language processes has increased greatly. One relevant factor was that new imaging techniques became available that allowed for visualizing brain activity on much more fine-grained temporal and spatial scales than was previously possible. The temporal dynamics of brain activation can be revealed millisecond by millisecond by neurophysiological imaging techniques such as EEG and MEG. The activation loci can be revealed using metabolic imaging techniques, positron emission tomogrophy (PET), and functional magnetic resonance imaging (fMRI). With fMRI, voxels of cortical volume as small as one cubic millimeter can be investigated. Thus, the tools became available to scrutinize the cortical mechanisms of cognition and language in space and time in the millimeter and millisecond ranges.

Some of the new studies confirmed views resulting from the classical neurological studies of aphasia. For example, the language areas of Broca and Wernicke were found to become active in various language tasks. Language comprehension experiments revealed strongest activation in the superior temporal lobe, and speech production tasks were found to elicit strongest activation in the inferior frontal cortex. Thus, basic predictions of the classic language model of Lichtheim could be confirmed (Price, 2000). It has also been suggested that the area most relevant for the processing of meanings and concepts associated with words is in the left inferior temporal lobe. As we will see later, the question about the processing of word meaning in the brain is a big challenge for modern neuroimaging research; it constitutes one of the most intensely debated topics in the neuroscience of language.

The advent of modern imaging research has also led to the generation of data about language-induced cortical activation that cannot be easily explained by the Lichtheim model and related classic neurological language theories. Broca's area, the "motor language center" of the Lichtheim scheme, became active in genuine language perception tasks such as phonetic judgment (Zatorre et al., 1992) or silent reading (Fiez & Petersen, 1998). This indicates that the function of the anterior "motor" language area is not restricted to language production, but includes aspects of language perception as well. Correspondingly, Wernicke's area, the "auditory language center," became active during genuine production tasks even when self-perception of the speech signal was made impossible by adding noise (Paus et al., 1996). Thus, the posterior "auditory" language area seems to have an additional role in language production. In conclusion, although the core language areas were active when language tasks had to be solved, their specific function

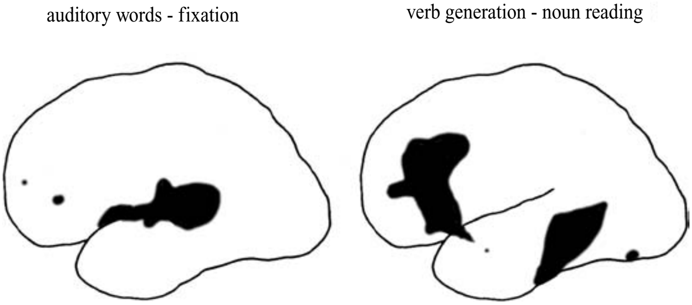

Figure 3.2. Presentation of spoken words led to increased metabolism in perisylvian areas relative to looking at a fixation cross (Left). In contrast, verb generation activated additional prefrontal areas and the middle and inferior temporal gyrus. Modified from Fiez, J. A., Raichle, M. E., Balota, D. A., Tallal, P., & Petersen, S. E. (1996). PET activation of posterior temporal regions during auditory word presentation and verb generation. *Cerebral Cortex*, *6*, 1–10.

as centers contributing to either speech production or comprehension became questionable. These neuroimaging results agree with neuropsychological findings indicating that both classical language areas are necessary for both language production and comprehension. Both main language areas appear to become active during and to be necessary for language processing, although neither of them is sufficient for word comprehension or production. Most likely, the two classical language centers play a role in both types of processes.

New experimental paradigms were introduced for imaging studies, because they were found to induce the most pronounced metabolic or neurophysiological changes. An example is the verb generation paradigm in which subjects are given nouns and must find verbs that refer to actions related to the referent of the noun. Examples would be “fork” and “eat,” or “horse” and “ride.” When subjects carried out this task, frontal areas including but not restricted to Broca’s area and temporal areas overlapping with or adjacent to Wernicke’s area were found to become active (Petersen et al., 1989; Wise et al., 1991).

Figure 3.2 shows the activation pattern in the left hemisphere obtained in one study in which verbs had to be generated in response to nouns (Fiez et al., 1996). When words had to be listened to, there was activity in superior temporal and inferior frontal areas that was absent when subjects had to fixate a cross. However, when new words had to be generated, there was much more widespread activation in the frontal and temporal lobes. This additional activity was also present when word generation was compared with reading of words. This result was used to argue that word processing can relate to the activation of various areas, not just to the classical language areas of Broca and Wernicke. Although this conclusion is compromised by difficulties in interpreting results from word generation experiments, it received support from additional neuroimaging work (for discussion, see Pulvermüller, 1999).

Which areas are relevant for processing which aspect of language? Clearly, if one must generate a word, this makes one think about possible candidates and alternatives and their respective adequacy, advantages, and disadvantages. If one must perform a more automatized task such as reading words or even reading very common words, this requires much less effort; consequently, some of the differences in brain activation between tasks can be accounted for easily in terms of the difficulty of the experiment. *Difficulty* is an ambiguous word here, and it may be better to speak about the degree of focused attention required; the amount of attention that must be shared between different cognitive entities (for example, a presented word and a to-be-generated word), the amount of effort that must be invested in a search (for example, for a new item), and numerous other cognitive processes that can become relevant in language-related tasks, including keeping items in memory, comparing items with each other, and rejecting items. Apart from these aspects, it is not surprising that the modalities involved in language tasks are related to the activation pattern. For example, a task in which visual stimuli are to be read and a task in which auditory stimuli must be processed may more strongly activate the superior temporal or posterior occipital lobe, respectively. A task involving language production may, compared to a mere comprehension task, induce more neuronal activity in action-related areas in the frontal lobe. Apart from such more or less trivial differences that can be expected to occur, and actually do occur, between language tasks, one may ask whether there are more specific correlates of language processing, for example, of the processing of the meaning of a word, its *semantics*.

The question concerning the cortical locus of the processing of word meaning has been addressed in many imaging studies, and a number of conclusions have been proposed. In a discussion in *Behavioral and Brain Sciences*, researchers involved in neuroimaging and neuropsychological research summarized their opinions about the cortical locus of semantic processes. Figure 3.3 presents the outcome of this discussion. Based on PET and EEG data, Posner and DiGirolamo (1999) argued for semantic processes in left inferior frontal areas. Salmelin, Helenius, and Kuukka (1999) reported MEG evidence that the left superior temporal lobe is relevant for word semantics. In Tranel and Damasio's (1999) framework, the role of the left inferior and middle temporal gyri was emphasized on the basis of

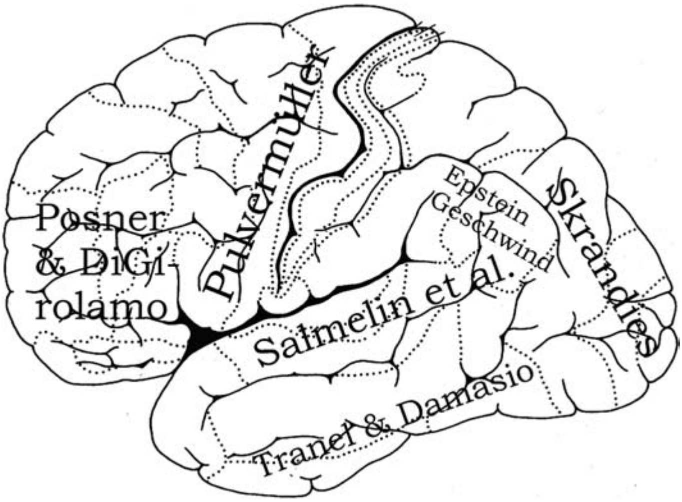

Figure 3.3. Cortical areas proposed to be related to the processing of word meaning. The name of the authors postulating the relevance of the respective areas is printed on top of Brodmann's area map. Note that different authors attribute the processing of word-related semantic information to different cortical areas. Modified from Pulvermüller, F. (1999). Words in the brain's language. *Behavioral and Brain Sciences*, *22*, 253–336.

both lesion evidence and PET data. Skrandies (1999) reported EEG studies suggesting that the occipital lobes might be involved in distinguishing word meanings. Epstein (1999) referred to the neuropsychological model by Geschwind (1965a, 1965b), according to which the angular gyrus at the boundary of the parietal, temporal, and occipital lobe relates to semantics. Finally, the opinion was expressed that action-related areas – primary motor, premotor, and prefrontal areas – also may be relevant for aspects of word-meaning processing and representation (Pulvermüller, 1999a). Regarding word semantics, Posner and DiGirolamo's (1999) statement "there is some dispute about the exact areas involved" is clearly correct. As this discussion shows, modern imaging work has not always led to a satisfactory answer of long-standing and important questions about the brain–language relationship. In spite of the undeniable progress made possible by empirical results obtained with newly introduced techniques, it is likely that theoretical advances are necessary as well. These may help to explain *why* the processing of meaning – and other language and cognitive processes – should activate certain cortical areas but not others, and may, in the best case, help us understand the diversity of results reported in some areas of the cognitive neuroscience of language. So far, it appears that many imaging studies suggest that numerous areas outside the classical language areas can become

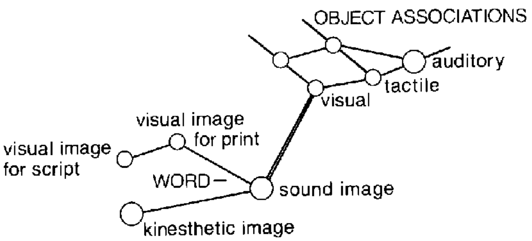

Figure 3.4. Freud's proposal of how a word is organized in the human brain. Multimodal associations are thought to exist between neuronal elements housed in different cortical systems. This model was the seed for modern theories of language processing in the brain (see Chapter 4). Adopted from Freud, S. (1891). *Zur Auffassung der Aphasien.* Leipzig, Wien: Franz Deuticke.

relevant for the processing of words and their meanings. This important issue is discussed further in Chapters 4 and 5.

Many areas outside the classical language areas of Broca and Wernicke are relevant for language processing. This idea is suggested not only by modern imaging research, but actually has a long history. In his monograph on aphasia, Freud (1891) proposed a neurological language model that competes with the one put forward by Lichtheim. Word representations are viewed as multimodal associative networks distributed over various cortical areas (Fig. 3.4). Unfortunately, this proposal was not well received over 100 years ago, a fact that forced Freud into a career outside the neurological domain. However, his ideas appear modern and attractive today. As shown in subsequent chapters (e.g., 4–6), models developed on their basis can explain important results from neuropsychological and neuroimaging research on language.

3.4 Summary

In the light of neuropsychological and neuroimaging research, Lichtheim's connectionist model (Fig. 3.1) was partly confirmed. However, it also became apparent that this model is too crude to explain many patterns of brain activation induced by language processing and the more fine-grained aspects of language disorders caused by brain lesions. It appears likely that the language centers of Broca and Wernicke are mutually functionally dependent. Furthermore, although these core language areas are certainly important for language, they do not appear to represent the only cortical areas contributing to and necessary for language processing. This is demonstrated by neuroimaging research showing that, in addition to the core language areas, other areas become active when specific language stimuli are being processed, and by converging neuropsychological reports about patients with lesions outside their core language areas who showed category-specific linguistic deficits. These studies demonstrate that, in addition to the *core language areas* of Broca and Wernicke, there are additional or *complementary language areas* that are activated during language processing whose lesion causes deterioration of aspects of language processing. These are not restricted to the language-dominant hemisphere (the left hemisphere in most right handers), but include areas in the nondominant right hemisphere as well. For the sake of clarity, it may appear advantageous to distinguish the core language areas of Broca and Wernicke in the language-dominant hemisphere from the complementary language areas in both hemispheres that can also play a role in aspects of language processing.

It may be possible to explain the mutual functional dependence of the core language areas and the category-specific role of complementary language areas. In Chapter 4, a neurobiological model is summarized that attempts at such explanation. The model uses concepts introduced in Chapter 2 – namely, the idea concerning functional cortical webs for words and other language elements. These are proposed to be cortically organized as strongly connected assemblies of neurons whose cortical distributions varies with word type. For example, concrete words referring to objects and actions are proposed to be organized as widely distributed cell assemblies composed of neurons in sensory and motor areas involved in processing the words' meanings (see Chapter 4). In contrast, highly abstract grammatical function words and grammatical affixes are assumed to be more focally represented in the left-hemispheric core language areas of Broca and Wernicke (see Chapter 6). This proposal builds upon the classic model, because neuron sets in core language areas are believed to be relevant for all types of language-related processes. However, it offers a dynamic perspective on the "concept area B" (Fig. 3.1) by proposing that complementary neurons in areas related to actions and perceptions regularly involved in language use contribute to category-specific language processes.

CHAPTER FOUR

Words in the Brain

This chapter complements Chapter 3 in providing neuroimaging and neuropsychological data about language. Here, the focus is on words. It is asked which brain areas become active during, and are relevant for, the processing of words in general, and that of specific word categories in particular.

An aim of this chapter is to show that the neuroscientific principles discussed in Chapters 2 and 3 give rise to new ideas about the representation and processing of words in the brain. The cortex, a neuroanatomically defined associative memory obeying the correlation learning principle, allows for the formation of distributed functional webs. During language acquisition, the neurobiological principles governing the cortex interact to yield the neuron machinery underlying language. Distributed functionally coupled neuronal assemblies, functional webs, are proposed to represent meaningful language units. These distributed but functionally coupled neuronal units are proposed to exhibit different topographies. Their cortical distribution is proposed to relate to word properties. It is asked how this idea fits into evidence collected with modern neuroimaging techniques.

4.1 Word-Form Webs

Early babbling and word production are likely caused by neuron activity in cortical areas in the inferior frontal lobe, including the inferior motor cortex and adjacent prefrontal areas. The articulations cause sounds, which activate neurons in the auditory system, including areas in the superior temporal lobe. The fiber bundles between the inferior frontal and superior temporal areas provide the substrate for associative learning between neurons controlling specific speech motor programs and neurons in the auditory cortical system stimulated by the self-produced language sounds. The correlation learning principle implies the formation of such specific associations resulting in functional webs distributed over the perisylvian cortex, which includes the inferior frontal and superior temporal core language areas. Figure 4.1a indicates the approximate left-hemispheric distribution of a functional web envisaged to realize a phonological word form. If neurons in the dominant left hemisphere are more likely to respond specifically to phonological features in the acoustic input, the resulting phonological networks must be lateralized, in the sense of having more neurons in one hemisphere than in the other. These lateralized perisylvian neuron ensembles would later provide the machinery necessary for activating a word's articulatory program as a consequence of acoustic stimulation with the same word form. This is necessary for the ability to repeat words spoken by others.

Interestingly, *babbling*, the infant's earliest language-like articulations, starts around the sixth month of life (Locke, 1989) and is immediately followed by the development of electrophysiological indicators of memory traces for phonemes (Cheour et al., 1998; Näätänen et al., 1997) and the infant's ability to repeat words spoken by others (Locke, 1993). These observations are consistent with the idea that babbling is essential for building up language-specific neuronal representations, in particular sensori-motor links that may in turn be essential for the ability of spoken word repetition. Word production, in the context of repetition or otherwise, may be essential for the build-up of specific neuronal representations of individual words.

It might be considered a shortcoming of this proposal that only a minority of word forms are learned by the infant based on single-word repetition (Pulvermüller, 1999b). How would infants know about which phonemes actually belong to one word or morpheme if it is spoken in continuous utterances of several words, with many word boundaries unmarked by acoustic cues? The answer is again implied by the correlation learning principle. The recurring sound sequences constituting words can be distinguished on statistical grounds from the more accidental sound sequences across word boundaries (Brent & Cartwright, 1996; Harris, 1955; Redlich, 1993). Behavioral evidence suggests that young infants distinguish the correlated phoneme and syllable sequences making up words from more accidental sound sequences in their acoustic input (Saffran, Aslin, & Newport, 1996). Therefore, it appears likely that single word input is not necessary to build up word representations, but that infants can use the correlation statistics, the transitional probabilities and/or mutual information of phoneme and syllable sequences (Shannon & Weaver, 1949), for learning words from continuous speech. After an auditory word representation has been established by correlation learning, the repeated articulation of the word made possible by the sensorimotor links set up by babbling would finally establish the word-related functional web.

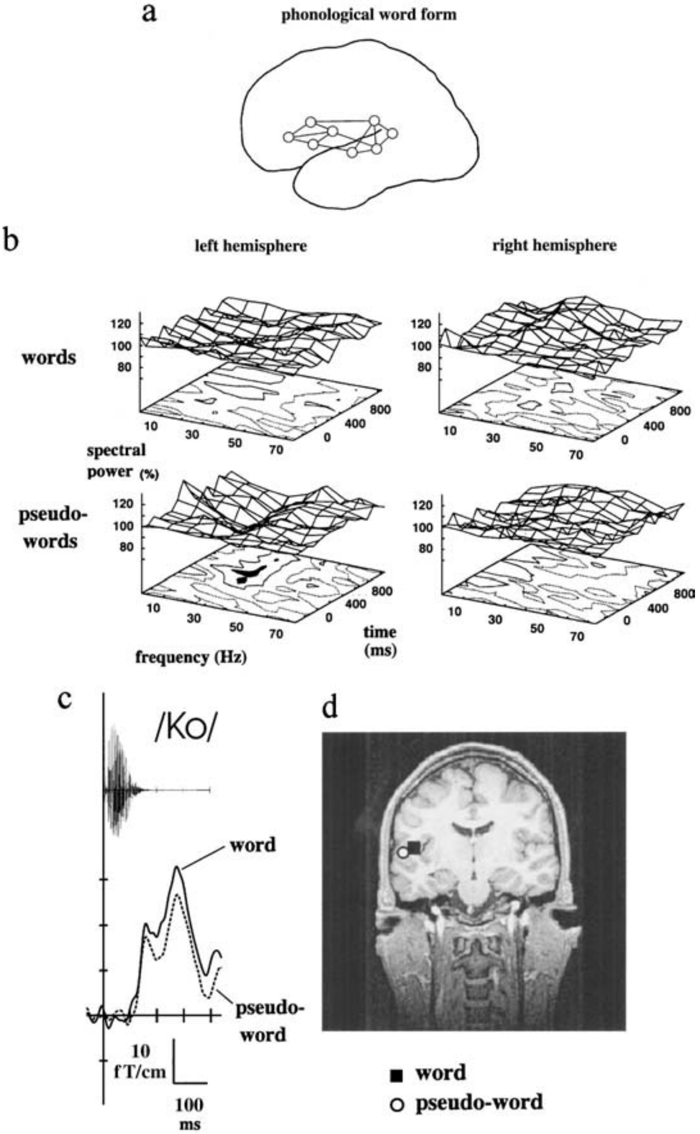

Figure 4.1. (a) The functional webs realizing phonological word forms may be distributed over the perisylvian areas of the dominant left hemisphere. Circles represent local neuron clusters and lines represent reciprocal connections between them. (b) Word presentation induced stronger gamma-band responses in the 30 Hz range compared to pseudo-word presentation, in particular over the left hemisphere. Reverberatory circuits within word webs may underlie the enhancement of high-frequency responses to words compared to pseudo-words. (c) The magnetic correlate of the Mismatch Negativity, the MMNm, was stronger in response to words compared to pseudo-words. Significant differences appeared already around 150 ms after the word recognition point, suggesting that the activation of word-related functional webs (lexical access) is an early process. (d) The main generator of the word-evoked magnetic mismatch response was localized in the left superior temporal lobe. Adopted from Pulvermüller, F. (2001). "Brain reflections of words and their meaning." *Trends in Cognitive Sciences,* 5, 517–24.

How would it be possible to prove the existence of functional webs relevant for the processing of words? One view on the nature of functional webs is that their massive reverberatory circuits produce precisely timed high-frequency rhythms when active (Pulvermüller, Birbaumer, Lutzenberger, & Mohr, 1997; Milner, 1974; Tallon-Baudry & Bertrand, 1999; von der Malsburg, 1995). Based on this assumption, the prediction is that words in the input activate the corresponding functional webs, thereby eliciting strong high-frequency rhythms. In contrast, phonologically and orthographically regular pseudo-words that are not part of the language would fail to activate a corresponding functional web, and the high-frequency activity in the perisylvian areas should therefore be relatively low.

This prediction was put to a test using MEG. In fact, a frequency band around 30 Hz revealed significant differences between acoustically presented words and pseudo-words. About one-half second after the onset of spoken one-syllable words, high-frequency brain responses were significantly stronger compared to the same interval following pseudo-words. Figure 4.1b shows the results of spectral analyses carried out on data recorded close to the left anterior perisylvian areas and the homotopic areas in the right hemisphere of a subject listening to words and pseudo-words. Word-induced high-frequency responses were markedly stronger compared to pseudo-word-induced activity, both in the single subject whose data are displayed (12 percent) and in the group average (8.4 percent) (Pulvermüller, Eulitz et al., 1996). This cannot be due to a global enhancement of the signal because event-related magnetic fields tended to be weaker for words than for pseudowords in the time window analyzed. EEG- and MEG-studies confirmed that words evoke stronger high-frequency brain activity than comparable wordlike material (Eulitz et al., 2000; Krause et al., 1998; Lutzenberger, Pulvermüller, & Birbaumer, 1994; Pulvermüller, Preissl, Lutzenberger, & Birbaumer, 1996). An explanation of these results can be based on the assumption that word presentation activates functional webs, including multiple reverberatory circuits, that fail to become fully active if pseudo-words are perceived.

Physiological differences between words and pseudo-words have been found in numerous studies using both electrophysiological and metabolic neuroimaging techniques (Creutzfeldt, Ojemann, & Lettich, 1989; Diesch, Biermann, & Luce, 1998; Hagoort et al., 1999; Price, Wise, & Frackowiak, 1996; Rugg, 1983). Thus, it is uncontroversial that the brain distinguishes between words and similar but novel and meaningless items. It is, however, unclear whether such physiological distinction would necessitate that experiment participants focus their attention on features of the stimuli or engage in language-related tasks. A further question is at which point in time, after the information about a word is present in the input, the brain makes the word–pseudoword distinction, so to speak. If distributed functional webs underlie word processing, an incoming verbal stimulus should automatically activate its corresponding representation. Given a sufficient number of input units specializing, for example, in the detection of acoustic stimulus features are activated, the entire strongly connected web would ignite automatically as a result of the strong feedforward and feedback connections holding the network together. This process of ignition (Braitenberg, 1978a) of the functional web should take place very rapidly, the major factor determining the latency being axonal conduction delays and temporal summation of neuronal activity. Axons can bridge large distances in the cortex within a few milliseconds, and the most common corticocortical fibers that have diameters of 0.5–1 μm can be estimated to propagate action potentials within 10–20 ms over distances of 10 cm (Aboitiz et al., 1992). A strongly connected distributed neuron web should therefore become active shortly after its initial stimulation, certainly within 100–200 ms, to use a conservative estimate.

Neurophysiological recordings, rather than the much slower metabolic neuroimaging techniques, are necessary to determine the point in time when the brain distinguishes words from pseudo-words. Some studies such as the studies of high-frequency cortical responses discussed previously have indicated that word-related brain processes can be detected late, that is, around 400 ms after the presence of the relevant information in the input. However, physiological word–pseudo-word differences in the event-related potential (ERP) of the brain have been found substantially earlier, in the so-called N1–P2 complex, 100–200 ms after onset of visually presented stimuli (Rugg, 1983).

In one series of EEG and MEG studies, we could confirm the early presence of neurophysiological indicators of word and pseudo-word processing. The Mismatch Negativity (MMN) and its magnetic equivalent (MMNm), which can be elicited by rare changes in the acoustic environment, were used. The MMN and MMNm were chosen because they have been found to reflect the existence of memory traces or engrams in the cortex and because they are largely independent of the subject's attention (Näätänen, 2001; Näätänen & Winkler, 1999). In earlier research, the MMN had been found to reflect the presence of memory traces for phonemes of the subjects' mother tongue (Näätänen, 2001). In a recent series of EEG and MEG studies, the neurophysiological correlates of spoken words were compared with the activity elicited by meaningless pseudo-words (Korpilahti et al., 2001; Pulvermüller, Kujala et al., 2001; Shtyrov & Pulvermüller, 2002b). To control for the physical difference, which necessarily distinguishes any word from a pseudo-word, two-syllabic items ending in the same second syllable were chosen. Between the two syllables was the pause characteristic of some Finnish consonants, so-called double-stop consonants (for example, “kk”). This pause made it possible to record separate non-overlapping brain responses to the two individual syllables of a naturally spoken bisyllabic word. To give an example, the Finnish word “pakko” meaning “compulsion” was contrasted to the meaningless pseudo-word “takko,” and physically identical “ko” syllables were used after the pause seperating the “pa” or “ta” on one side and the critical syllable “ko” on the other. Thus, the same critical syllable was placed in a word or pseudo-word context, respectively.

When the critical second syllable completed a word, its MMN and MMNm were larger compared to when the syllable was placed in a pseudoword context (Fig. 4.1c). The respective difference was most pronounced 100–200 ms after the word recognition point of the lexical items. The *word recognition point* is the earliest point in time when the information present in the acoustic input allows the subject to identify the word with some confidence (Marslen-Wilson & Tyler, 1980). This suggests that the functional web activated by a word in the input becomes active quite early. This finding is also consistent with recent results from dynamic statistical parametric mapping based on fMRI and MEG (Dale et al., 2000). These results indicate access of the cortical representation of words within the first 200 ms after stimulus information is present in the input. (Pulvermüller, Kujala et al., 2001).

The main source of the cortical generator of the MMNm was localized in the left superior temporal lobe (Figure 4.1d). Whereas this source was stronger for words than pseudo-words, its anatomical locus did not change with lexical status.

It thus appears that the brain can distinguish words from pseudo-words quite early after the relevant information is present in the input. Still, other physiological word–pseudoword differences – in particular, in high-frequency activity – tend to occur with longer latencies. This may, in part, be related to differences between experiments (regarding stimuli, tasks, etc.), but may as well indicate that different brain processes are reflected by these measures. Early ERP differences may reflect the initial full activation, ignition (Braitenberg, 1978a) of memory traces for words, a putative correlate of word recognition, whereas differences in high-frequency responses may reflect continuous reverberatory activity of word-related functional networks, a putative state of active memory (Fuster, 1997).

It is noteworthy that in the studies of the MMN and MMNm elicited by words (Korpilahti et al., 2001; Pulvermüller, Kujala et al., 2001; Shtyrov & Pulvermüller, 2002a, 2002b), the early enhancement of these responses to words was seen, although experiment participants were instructed to direct their attention away from the acoustic input and watch a silent movie. Together with results from metabolic imaging studies (Price et al., 1996), the physiological distinction of words and pseudo-words in these experiments proves that attention to words is not necessary for activating the words' cortical memory traces.

In summary, physiological studies provide support for the existence of word representations in the brain. The enhanced high-frequency responses in the gamma band to words are consistent with coordinated fast reverberatory neuronal activity generated by functional webs.

4.2 Category-Specific Word Webs

Word use in the context of objects and actions leads to associations between neurons in the cortical core language areas and additional neurons in areas processing information about the words' referents. This is implied by the correlation learning principle and the cortex's long-range connections between motor and sensory systems. Functional webs could therefore provide the basis for the association, in the psychological sense, between an animal name and the visual image it relates to, or between an action verb and the action it normally expresses. Strong links within the web can account for one's impression that the image is automatically aroused by the word form presented alone, and that, vice versa, the image almost automatically calls the name into active memory. The neuron ensembles linking phonological information and information about the actions and perceptions to which a word refers are termed *word webs* here. They would include the phonological webs in perisylvian areas and, in addition, neurons in more widespread cortical areas critically involved in processing perceptions and actions, and, additional neurons in various cortical sites where sensory and action-related information converges and is being merged. The type of entity a word usually refers to should be reflected in the cortical topography of the functional web that realizes it.

4.2.1 Visually Related and Action Words

The meaning of an animal name, such as "whale" or "shark," is usually known from visual experiences, pictures, or films, whereas the meaning of a tool name, such as "nail" or "fork," refers to objects one uses for certain actions. This is not to say that one could not know a whale from interacting with it or forks from looking at them, but it may appear plausible that in most individuals more relevant information characterizing whales and forks is visually or action related, respectively. In principle, to draw firm conclusions on perceptual and functional attributes of word and conceptual categories, the perceptual and action associations of the stimuli must be evaluated empirically. The lack of such stimulus evaluation is a caveat of many studies of category-specific brain processes. Behavioral investigations carried out with healthy volunteers revealed that many animal and tool names show the expected differential elicitation of visual or action associations, respectively. However, the most striking double dissociation in perceptual and action attributes was seen between action verbs on the one hand and selected nouns referring to animals or large human-made objects on the other (Fig. 4.2d) (Pulvermüller, Lutzenberger, & Preissl, 1999). In addition, categories such as "animal names" were not well defined with regard to the modality for which most striking associations were being reported. For example, whereas words such as "whale" or "shark" are reported to elicit primarily visual associations, the results for "cat" are less clearcut, for obvious reasons. Thus, the differential associations cut across the categories suggested by a philosophical approach (e.g., living vs. nonliving), as was earlier found for category-specific neuropsychological deficits (Warrington & McCarthy, 1987). The sensory/action modalities through which the referent of a word is known appear to be relevant (Fuster, 1999).

It has been argued that it is a possible limitation of this line of thought that it can be applied only to communication, where words are being learned in the context of their referent objects or actions. However, word meanings can also be picked up from contexts in which the actual referents are absent. Their meaning is frequently revealed by other words used in the same sentence or piece of discourse. It has been proposed that word meaning can be defined in terms of the set of other words that frequently co-occur with a given word (Landauer & Dumais, 1997). This would translate into a different neurobiological scenario for the learning of word meaning. Given that there is a stock of words whose meaning has been acquired based on word–object or word–action contingencies, a new word occurring in good correlation with such known words would only activate its phonological perisylvian representation, because no semantic (reference related) links have been set up. However, neurons in extraperisylvian space related to the known meaning of common context words would frequently be active together with the phonological web of the new word. The correlated activity of semantically related neurons included in the neuronal representations of known context words and the phonological web of the new word may allow for "piggy bay" learning of word meaning. Clearly, this implies that the semantically related neurons will finally be shared between previously known and new

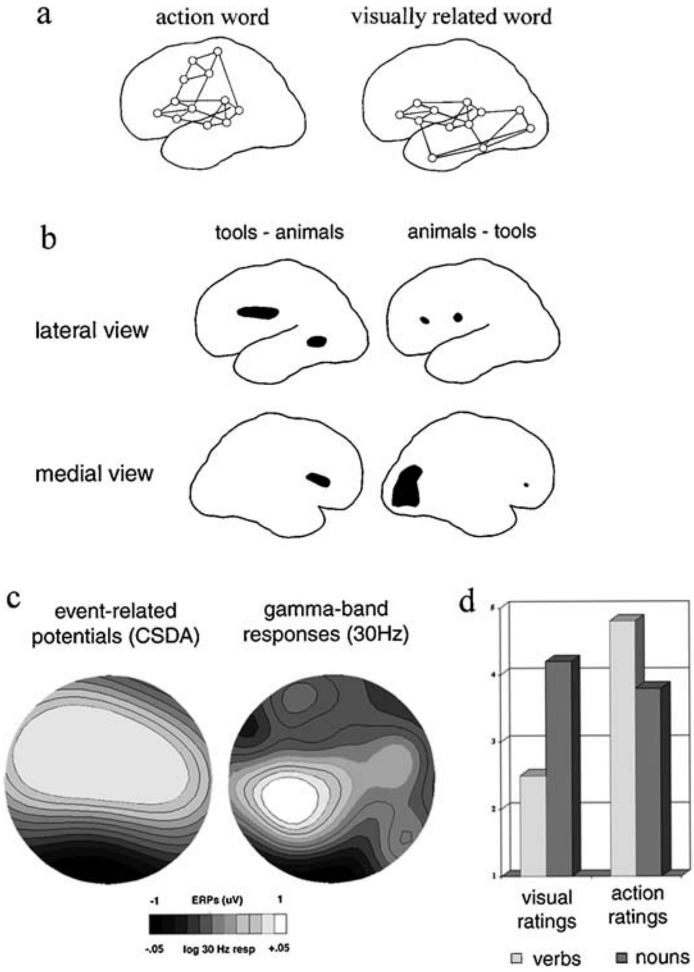

Figure 4.2. (a) Visual and action associations of words may be mapped by functional webs extending over perisylvian language areas and additional visual- and action-related areas in the temporo-occipital and fronto-central areas. The cortical topography of word-related functional webs of words primarily characterized by visual associations may therefore differ from those of words with strong action associations. (b) Differences in metabolic brain activation related to the processing of nouns referring to animals and tools in a naming task. Whereas the tool words more strongly activated a premotor region and an area in the middle temporal gyrus, animal names most strongly aroused occipital areas. (c) Electrophysiological differences between nouns and verbs in a lexical decision task recorded at central (close to motor cortex) and posterior (above visual cortex) recording sites. Gamma-band responses in the 30 Hz range were stronger close to the motor cortex for action verbs, and stronger above visual areas for nouns with strong visual associations. A similar difference was revealed by event-related potentials submitted to Current Source Density Analysis (CSDA). (d) Behavioral experiments showed that the stimulus nouns elicited strong visual associations whereas the verbs were primarily action related. Adopted from Pulvermüller, F. (2001). "Brain reflections of words and their meaning." *Trends in Cognitive Sciences,* 5, 517–24.

words so that their neuronal representations would overlap in their semantic parts. This line of thought shows that the learning of word meaning based on correlated neuronal activity is, by no means, restricted to the word–object contingency scenario. Word meaning can also be learned from context.

Figure 4.2a sketches the postulated neuronal architectures of functional webs representing words with strong visual vs. action associations, respectively. The circles in the diagrams symbolize local clusters of neurons strongly linked by corticocortical fibers. The diagrams illustrate the idea of word webs that include neurons reflecting semantic aspects. More precisely, the proposal is that cortical neurons processing aspects of the words' typical referents, the entities they refer to, are woven into the networks. If the referent is an object usually perceived through the visual modality, neurons in temporo-occipital areas should be included in the web. If a word refers to actions or objects that are being manipulated frequently, neurons in fronto-central action-related areas are assumed to be wired into the cortical representations. This can easily be extended to other sensory modalities as well (Pulvermüller, 1999b). A shortcoming of the diagrams is that only one type of association is shown for each word web. Usually, a word that is primarily visually related is reported to elicit some action associations as well; conversely, an action-related word yields some visual associations (cf. Fig. 4.2d). The all-or-nothing aspect one may infer from the diagrams is therefore unrealistic. To make the diagrams more realistic, each web should include some additional neurons in the respective other field, although the ensembles' neuron density in these additional areas would be lower than in the areas processing the words' primary referential aspect (Kiefer & Spitzer, 2001; Pulvermüller, 2001). Furthermore, action associations imply that there are associations to self-perceived aspects of the action in the somator-sensory or visual modality. The visual perception of one's own hand during knocking or writing likely arouses neurons in movement-related areas of the visual system not activated if, for example, a stationary visual stimulus is perceived. Therefore, there would be good reason to add detail to the diagrams, however, at the cost of making them more complex. Nevertheless, the topographies of functional webs may differ between semantic word types, and the diagrams may convey the gist of this idea. (Fig. 4.2a, Fig. 4.3a).

The postulated differential topographies of word webs imply meaning-related processing differences between word categories. A major source of evidence for such differences is neuropsychological patient studies in which, for example, the production or comprehension of nouns and verbs or animal and tool names was found to be affected differentially by disease of the brain (Bak et al., 2001; Cappa et al., 1998; Daniele et al., 1994; Miceli, Silveri, Nocentini, & Caramazza, 1988; Miceli, Silveri, Villa, & Caramazza, 1984;

Patterson & Hodges, 2001; Warrington & McCarthy, 1983; Warrington & Shallice, 1984; see also Chapter 3). These dissociations between kinds of words and conceptual categories can be understood on the basis of the assumption of distributed neuron ensembles reflecting perceptual and structural attributes, including visual features and the degree of overlap between exemplars, and the functional attributes, the actions to which the words and concepts relate (Humphreys & Forde, 2001).

It also may be asked whether the intact human brain demonstrates differential activation of brain areas when action- or perceptually related words are being processed. One example of a critical prediction is, if words of one kind are characterized by stronger action (visual) associations than are words of another kind, their processing should be accompanied by stronger brain activity in the relevant action- (sensory-) related areas. Relevant action-related areas are in the frontal lobe; the areas necessary for object perception are in the occipital and inferior temporal lobes.

When pictures of animals and tools were presented in a naming experiment, several areas, including occipital and temporal sites and the classical language areas, were found to increase their activity (Martin et al., 1996). Category-specific activation was found in the premotor cortex and the middle temporal gyrus when tools had to be named silently, and in the occipital and inferior temporal lobe when animals had to be named (Fig. 4.2b). These results meet the previously noted predictions. One may speculate that the premotor activation is related to the action associations of tool names, as the activation in inferior–temporal and occipital areas may be related to visual attributes of animal names. The additional activation in the middle temporal gyrus in tool naming may be related to movement associations elicited by the words involved. Differential cortical activation by action- and visually related concepts and words were confirmed, in part, by more recent metabolic imaging studies of category-specific processes using PET and fMRI (Damasio et al., 1996; Grabowski, Damasio, & Damasio, 1998; Martin & Chao, 2001; Moore & Price, 1999; Mummery et al., 1998; Noppeney & Price, 2002; Perani et al., 1999; Spitzer et al., 1998; Warburton et al., 1996), although not all researcher could confirm such differences (e.g., Devlin et al., 2002).

Neurophysiological investigation of noun and verb processing provided further evidence for category-specific brain processes relevant for language (Brown & Lehmann, 1979; Koenig & Lehmann, 1996; Molfese, Burger-Judisch, Gill, Golinkoff, & Hirsch-Pasek, 1996; Preissl, Pulvermüller, Lutzenberger, & Birbaumer, 1995; Pulvermüller, Lutzenberger et al., 1999; Pulvermüller, Preissl et al., 1996; Federmeier et al., 2000). In one study (Pulvermüller, Lutzenberger et al., 1999), differential visual and action associations of the nouns and verbs selected were demonstrated by a rating study performed by all experiment participants. Event-related potentials (ERPs) and high-frequency cortical responses revealed a physiological double dissociation consistent with differential activation of fronto-central and occipital areas (Fig. 4.2c). The ERP difference was apparent approximately 200 ms after the onset of visual word stimuli, consistent with early activation of the word webs involved. Topographically specific high-frequency responses, which were stronger over central areas for verbs and over occipital areas for nouns, started later (400 ms). In an EEG study of auditory word processing, the physiological distinction between visually related nouns and action verbs could be replicated, and similar differential activation was found between visually and action-related nouns. In contrast, there was no difference in the topography of brain responses between action verbs and nouns for whom strong action associations were reported (Pulvermüller, Mohr, & Schleichert, 1999). These topographical differences in the activation patterns elicited by action-related and visually related words resemble the differences observed between written tool and animal names, and between pictures of animals and tools (Kiefer, 2001). All of these results indicate that the differential activity pattern evoked by word kinds is not grammatically related, but rather reflects semantic properties of the stimulus words and their related concepts. Pulvermüller, Assadollahi, and Elbert (2001) found a global enhancement of the evoked brain response for a certain subcategory of nouns, which, according to their behavioral data, had particularly strong semantic associations to both objects and actions (multimodal semantics). Control nouns without multimodel semantics failed to elicit the result, again arguing against an interpretation in terms of grammatical word categories. Differences in word semantics may also underlie the neurophysiological differences found between lexically ambiguous and unambiguous words (Federmeier et al., 2000). This interpretation is suggested because it may appear plausible that the meaning of ambiguous words is somewhat richer than that of words with only one meaning. When semantic properties of the stimulus words were systematically evaluated, we found a linear increase of an early component of the event-related magnetic field with a measure of the strength of semantic associations of a word ($r = 0.8$). Therefore, these data, along with those mentioned earlier, enforce an account in terms of word semantics. It may be that the strong associations, in the psychological sense, of words with mutimodal semantics are realized as strong connections within particularly widespread and large cortical neuronal assemblies. Activation of these particularly widespread and strongly connected networks may be reflected as an enhancement of the neurophysiological response (Pulvermüller, 2001).

4.2.2 Sub-types of Action Words

More fine-grained predictions are possible on the basis of the postulate that topographies of word webs reflect the words' referents. Action verbs can refer to actions performed with the legs (walking), arms (waving), or mouth (talking). It is well known that the motor cortex is organized somatotopically, that is, adjacent body muscles are represented in neighboring areas within the motor cortex (He et al., 1993; Penfield & Rassmussen, 1950). Neurons controlling face movements are located in the inferior precentral gyrus, those involved in hand and arm movements accumulate in its middle part, and leg movements are controlled by neurons in its dorsomedial portion (see Chapter 2). The correlation learning principle therefore suggests differential topographies for cell assemblies organizing leg-, arm-, and face-related words (Fig. 4.3a). Differential action-related associations of subcategories of verbs could be demonstrated by behavioral studies (Fig. 4.3b; Pulvermüller, Hummel, & Härle, 2001).

In an EEG study, we compared face- and leg-related action verbs ("walking" vs. "talking"). Current source density maps revealed differential activation along the motor strip. Words of the "walking" type evoked stronger ingoing currents at dorsal sites, over the cortical leg area, whereas those of the "talking" type elicited the stronger currents at inferior sites, next to the motor representation of the face and articulators (Fig. 4.3c; Hauk & Pulvermüller, 2002; Pulvermüller, Härle, & Hummel, 2000; Pulvermüller, Hummel, & Härle, 2001). A similar study comparing arm- and leg-related words was performed using fMRI. The preliminary data shown in Figure 4.3d are consistent with the view that the body parts involved in the actions action verbs refer to are reflected in the cortical neuron webs these words activate. Furthermore, the early point in time at which the word category differences were present in neurophysiological responses indicates that there was no substantial delay between word form access (perisylvian activation) and the processing of action attributes (more superior activation, for example in the case of leg words). This supports the view that information about the word form and the body parts with which the word-related actions are being carried out, are woven into the same word-related cortical networks and are activated near-simultaneously (Pulvermüller, 2001).

4.3 The Time Course of Lexical and Semantic Activation

The lexical status of a written or spoken word, whether it is a word or not, and aspects of word semantics appear to crucially determine the brain response. The differences between semantic word categories can appear early in the neurophysiological brain response, that is, ~100–200 ms after stimulus onset

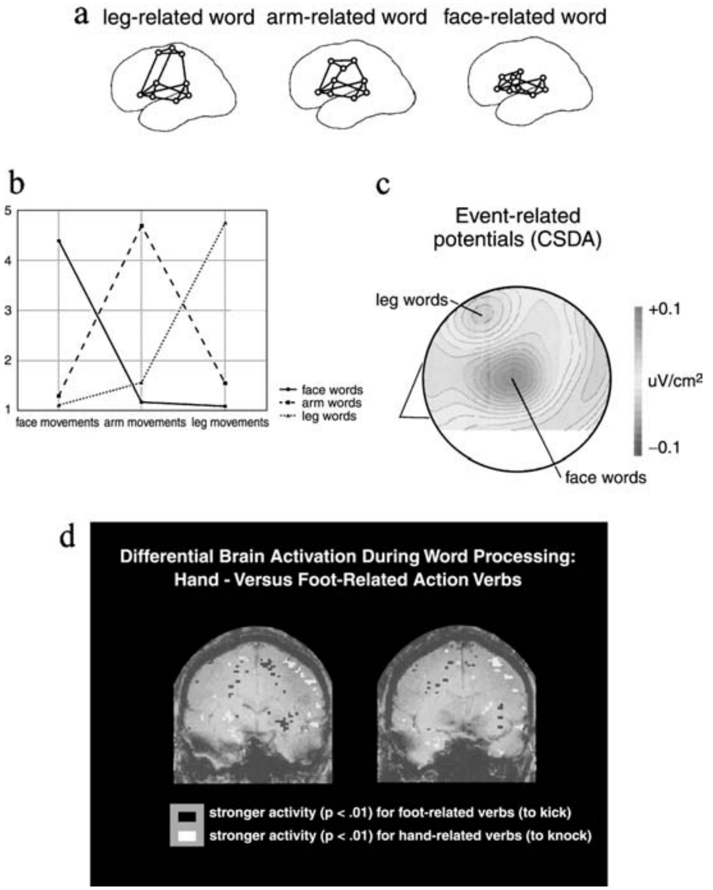

Figure 4.3. (a) Cortical topographies of functional webs representing different types of action words may differ. Action words can refer to actions executed by contracting face, arm, or leg muscles (e.g., *to lick, to pick, to kick*). Different neuron ensembles in the primary motor cortex may therefore be woven into the word-related neuron ensembles (cf. Fig. 1b). (b) Ratings of face, arm, and leg associations confirming differential referential semantics of three action verb groups. (c) Results from an EEG study. Topographical differences between brain responses to face- and leg-related verbs. Stronger ingoing currents were seen close to the vertex for leg-related items (gray spot at the top) and at left-lateral sites, close to the face representation, for face-related words (dark spot in the middle). (d) Result from a fMRI study comparing arm- and leg-related verbs (single subject data). Differences were see in the precentral gyrus of the left hemisphere. Adopted from Pulvermüller, F. (2001). "Brain reflections of words and their meaning." *Trends in Cognitive Sciences*, 5, 517–24.

(e.g., Pulvermüller, Assadollahi et al., 2001; Skrandies, 1998). This latency range corresponds to the time range where the earliest neurophysiological differences between words and pseudo-words were found (e.g., Pulvermüller, Kujala et al., 2001; Rugg, 1983). Thus, the earliest latencies at which the lexical status and the semantic category of word stimuli were reflected in the neurophysiological response coincided with each other. These neurophysiological data support psycholinguistic models postulating that information about a word's meaning can be accessed near-simultaneously with information about its form, a proposal motivated by behavioral studies (Marslen-Wilson & Tyler, 1975; Marslen-Wilson & Tyler, 1980). Likewise, they are consistent with the view that a word is cortically processed by a discrete functional unit storing information about the word's form together with that about its semantics.

While the semantically and form-related parts of distributed word webs may be activated early and near-simultaneously, there is evidence that different physiological processes occur in sequence in the same cognitive brain representations. A stage of access to the representation (ignition of the cell assembly, see Braitenberg, 1978a) may be followed by sustained reverberatory activity (active memory, see Fuster, 1995) of the word web. Whereas the early access stage may occur within one quarter of a second after the information in the input allows for recognizing a stimulus word, the reverberatory activity related to active memory would follow after more than 250 ms. The early access process may be reflected in early event-related potentials, and the late reverberations may lead to high-frequency responses in the gamma band. These hypotheses can tentatively explain recent findings about the time course of neurophysiological responses to words (for further discussion, see Kiefer, 2001; Pulvermüller, 1999).

4.4 Summary and Conclusions

The brain response to words and word-like material appears to reflect lexical status and word semantics. Word/pseudo-word and word category differences were reported in metabolic and neurophysiological imaging studies. Both types of differences were found already at 100–200 ms after the information in the input allowed for recognizing the words, whereas some differences, for example in high-frequency responses, appeared only with longer delays. These results can be explained on the basis of the idea that words are represented and processed by distributed but discrete neuron webs with distinct cortical topographies. They are somewhat less easily explained by alternative approaches. If words were represented by single neurons, for example, the corresponding specific brain activity states could probably not be distinguished with large-scale neuroimaging techniques, such as MEG or fMRI. Also, it is in question how the specific changes observed between words and pseudowords could be explained if both stimulus types were processed alike by a distributed network of neurons in which no discrete representations exist, or by interference patterns over the entire cortex. Furthermore, an explanation of word-category differences may turn out to be even more difficult on the basis of such approaches. Thus, while competing approaches are challenged by the data discussed, the postulate of discrete functional webs representing words explains them well.

The results on category differences described here indicate that aspects of the meaning of words are reflected in the topography of brain activation. They are also consistent with the view that the referents of particular word kinds are relevant for determining the brain areas involved in their processing. The data do not explain the entire spectrum of areas found to be active during category-specific word processing. There are findings about different semantically related activity patterns that are not readily explained by elementary neuroscientific principles, such as the principles (I) – (IV) discussed earlier. For example, the differential activation of the right- vs. the left parietal areas by names of body parts and numerals (Le Clec'H et al., 2000) cannot be explained by the four principles alone. It is likely that additional principles of cortical functioning as yet not fully understood are necessary to account for these data. Furthermore, it must be added that the semantic category of the stimulus words is by far not the only variable determining the topography of brain activation. Clearly, the modality of stimulation (visual or auditory) and the task context in which words must be processed (e.g., lexical decision, naming, memory) play an additional important role in determining the set of active brain areas (Angrilli et al., 2000; Mummery et al., 1998). Moreover, other features of the stimulus material, for example the length and frequency of words, play an important role (Assadollahi & Pulvermüller, 2001; Rugg, 1990). The present approach suggests, and the summarized data indicate, that, if possibly confounding variables are appropriately controlled for, category-specific differences are present between word categories and conceptual kinds across different tasks and stimulus modalities.

A limitation of the considerations so far is that they are restricted to the level of single words. In fact, they apply to word stems that include only one meaningful unit (morpheme). The neurobiological nature of affixes with grammatical function, for example, has not been addressed. Also, the neurobiological basis of grammatical or function words, for example "if", "is", and "it", has not been specified, and, of course, the complex interaction between word-related processes during analysis and synthesis of a word chain is a further open issue. Some of these issues will be covered in later chapters. Chapter 5 will address meaning and form relationships between words, Chapter 6 function words and inflectional affixes, and Chapters 10 to 13 various grammatical phenomena.

EXCURSUS ONE

Explaining Neuropsychological Double Dissociations

The cortex may be an associative memory in which correlation learning establishes discrete distributed functional webs. It is important to ask how this view relates to clinical observations, in particular to the *neuropsychological double dissociations* seen in aphasic patients. Here is an example of such a double dissociation. Patient A exhibits severe deficits in producing oral language (Task 1), but much less difficulty in understanding oral language (Task 2). Patient B, however, presents with the opposite pattern of deficits – that is, only relatively mild language production deficits, but substantial difficulty in comprehending spoken language. Briefly, Patient A is more impaired on Task 1 than Task 2, whereas Patient B is more impaired on Task 2 than Task 1.

Clearly, the explanation of neuropsychological syndromes and, in particular, of double dissociations is an important issue for any model of brain function and, therefore, a brief excursus may be appropriate here. It was once argued that the existence of a double dissociation demonstrates, or strongly suggests, the presence of modules differentially contributing to specific aspects of the tasks involved (Shallice, 1988). A module would be conceptualized as a largely autonomous information processor (see also Section 6.2). A standard explanation of a double dissociation, therefore, is that two modules are differentially involved in the two tasks (1 and 2), and that one of them is selectively damaged in each of the patients (A and B).

The idea of functional cortical webs may lead the reader to conclude that the distributed neuronal systems would always deteriorate as a whole when a lesion takes place, even if only a focal cortical area is affected. The suspicion may therefore be that double dissociations cannot be explained. This conclusion is, however, not justified. Neuropsychological double dissociations can, in fact, be accounted for by a model composed of distributed functional webs.

This excursus explains how a functional web architecture can model a double dissociation, and why double dissociations do not prove the existence of modules differentially involved in the tasks on which a double dissociation is observed. The initial paragraphs present considerations and simulations regarding the influence of small lesions on the function of an ensemble of strongly connected neurons.

E1.1 Functional Changes in a Lesioned Network: The Nonlinear Deterioration of Performance with Growing Lesion Size

How does a brain lesion change the function of a strongly connected neuronal ensemble? If one neuron in an assembly is lesioned, this may have no effect at all. However, if a certain percentage of its neurons have been lesioned or removed, the ensemble becomes unable to serve its functional role in cerebral life, so that it becomes inappropriate to speak about "full" activation or ignition when the remaining neurons become active. The smallest percentage of active neurons (of the intact assembly) necessary for speaking about an ignition is called the *ignition threshold* here. In the intact network, the ignition threshold can be considered the point of no return, the number of neurons active after which the ignition can no longer be stopped by inhibitory input. If the number of assembly neurons surviving a lesion is smaller than the ignition threshold, the ensemble, by definition, cannot ignite after stimulation. Therefore, it can be considered to be destroyed. Clearly, not every lesion leads to destruction of all affected ensembles. If the damage is moderate, the assemblies still ignite after appropriate stimulation, but the time needed for the ignition to take place is longer. This can be illustrated by neural network simulations (Pulvermüller & Preissl, 1991).

Figure E1.1 gives average ignition times t_θ and percentage of destroyed neuronal ensembles d as a function of the lesion size l. In this simulation, the ignition threshold has been set to 70 percent of the total number of neurons of the intact ensembles. Importantly, the assemblies tolerate lesions of a substantial percentage of their neurons with only minor ignition delays. Even pronounced lesions (e.g., of 15 percent of the neurons) affect performance only minimally. However, after a certain critical amount of damaged tissue has been reached, 20 percent in the presented simulations, performance deteriorates very rapidly if the lesion size increases slightly.

The nonlinear decline of performance of the networks with increasing lesion size may have implications for the explanation of progressive neurocognitive impairments. Many progressive brain diseases lead to minor behavioral deficits, although pronounced structural deficits can already be observed. Minor additional structural deterioration sometimes leads to rapid

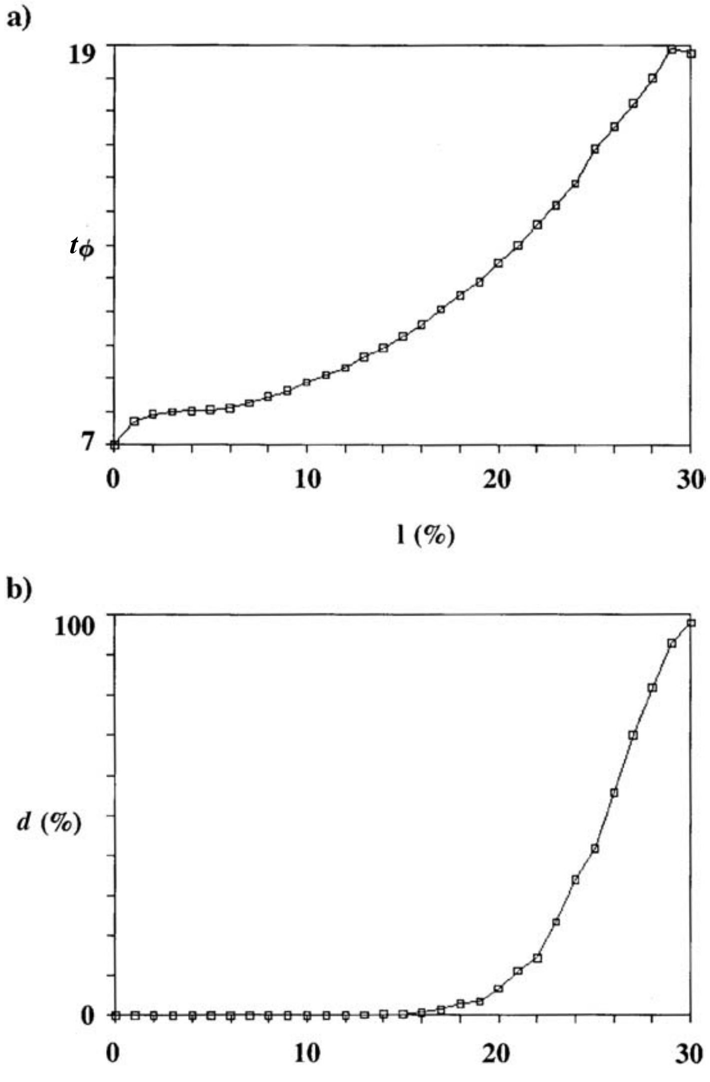

Figure E1.1. The effects of lesions (l) of different sizes on neuron ensembles each including 100 neurons. The ignition threshold was set to 70% – that is, the assembly was called *ignited* if 70 of the 100 neurons were active. As a function of lesion size, the average time t_θ needed for an ignition, and the percentage *d* of destroyed assemblies (which could no longer ignite) increased. Small lesions did not have a strong effect. In contrast, after removal of approximately 20% of the neurons, further slight increase of the lesion size caused dramatic dysfunction, as reflected by the steep raise of both ignition times $t_\emptyset$ and the percentage of destroyed ensembles *d*. Reprinted with permission from Pulvermüller, F., & Preissl, H. (1991). A cell assembly model of language. *Network: Computation in Neural Systems, 2*, 455–68.

and pronounced deterioration of behavior. The simulations show that this is exactly the way a functional web acts when progressively more of its neurons are being damaged. Small increases of the lesion size cause pronounced functional impairments only if ignition thresholds are being approached.

E1.2 Broca's vs. Wernicke's Aphasias

What is the effect of focal lesions in different parts of the perisylvian language cortex in a model of overlapping functional webs distributed over these cortical areas? Again, the idea is that acoustic, articulatory, and semantic information about words are bound together in functional units exhibiting specific cortical topographies. This means that these aspects of information processing are not separate functionally, although, originally, they may primarily have been housed in separate brain areas. How would a network of several distributed and partly overlapping information processors respond to focal lesions?

Before associative learning, articulatory programs are controlled by neurons in the prefrontal, premotor, and primary motor cortex, whereas acoustic properties may activate neurons only in the superior temporal lobe stimulated by features of speech sounds, and input related to word semantics (reference) are possibly exclusively processed in additional brain areas. After word learning, however, all (or most) of these neurons would be activated during comprehension, articulation, and semantic processing of a word because the functional web has formed. In the same way the neurons in different areas were functionally separate before learning had taken place, these distant neuron groups may become functionally separate again after their strong links have been cut by a lesion, or after one part of the assembly has been destroyed. If processes dissociate after the lesion, they may nevertheless have been linked in the fully functionl brain.

An intact neuronal assembly would include efferent neurons that control articulatory movements and afferent neurons stimulated by acoustic phonological input. These neuron groups can be considered to lie in the periphery of the assembly. In the center are neurons in various association cortices that do not receive direct input from outside the cortex and do not control its output directly. Their primary purpose is to bind information. These binding sites can themselves be connected to other areas that do not receive direct input from outside the cortex, or directly control motor output. Figure E1.2 schematizes a network of several partly overlapping neuronal ensembles and the way their elements (local neuron clusters) may be distributed over perisylvian cortices. The network equivalents of areas in the temporal lobe are represented at the top and those of the inferior frontal

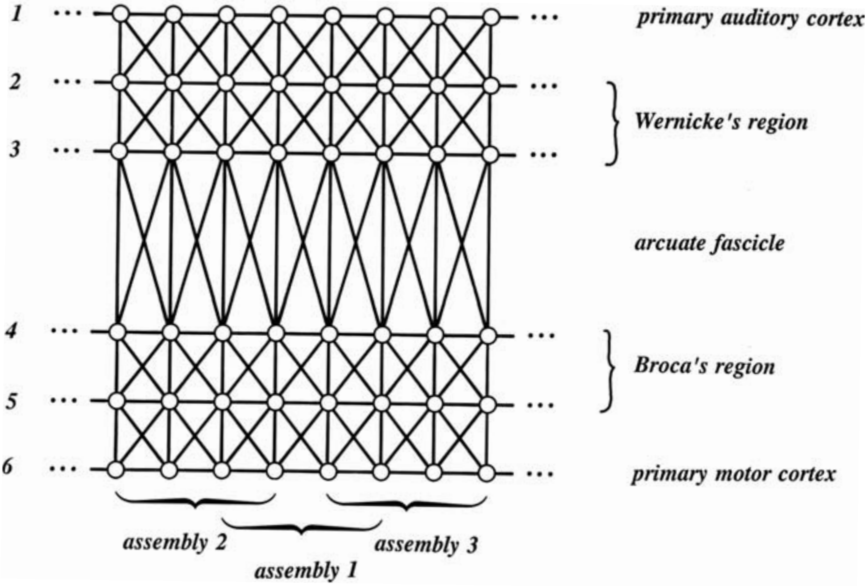

Figure E1.2. Structure of a network used for simulating the effect of focal lesions in the perisylvian cortex. Assemblies included neurons in primary and higher-order perisylvian areas. Lesions in one of the "peripheral" parts of the assemblies (uppermost input or lowermost output layers) led to unimodal processing deficits in the simulation (in either "word production" or "word perception"). Lesions in the middle (the network equivalent of Broca's and Wernicke's areas) caused multimodal but asymmetric deficits (e.g., a strong "production" deficit with a moderate "comprehension" deficit, or vice versa). The model therefore accounts for a double dissociation of the performance on two language tasks (speech production vs. comprehension), and, in addition, for the frequent cooccurrence of deficits regarding these tasks. Reprinted with permission from Pulvermüller, F., & Preissl, H. (1991). A cell assembly model of language. *Network: Computation in Neural Systems, 2*, 455–68.

cortex are at the bottom. Only the uppermost and lowermost neurons in the periphery (uppermost or lowermost layers) have efferent or afferent cortical connections. Obviously, this is an elementary model based on several simplifying assumptions.

Word comprehension is modeled in the circuit as stimulation of the upper auditory input layer and subsequent ignition of one assembly. Word production is modeled by stimulation in the center of one assembly followed by its ignition plus activation of its motor output neurons in the lowermost layer. Specific deficits in comprehension or production can now be explained intuitively as follows: A lesion in one of the layers or areas of the model destroys assembly neurons. However, lesions in the periphery, in a primary area, lead to additional disconnection of the web from its input or output. A moderate lesion in the uppermost layer may therefore only cause a mild increase of ignition times in the simulation of word production, thus leaving word production largely 'intact.' However, this same lesion may make it impossible to ignite the assembly through its afferent fibers if word comprehension is simulated. This impossibility does not necessarily imply that all the assembly's afferent connections have been cut or that all neurons in the auditory input layer have been destroyed. Removal of only a few peripheral input neurons of an assembly can slightly delay its ignition so that its neighbors may take over and ignite instead of the directly stimulated assembly. This process corresponds to a failure or error in activating a word-specific neuron ensemble. One may consider this an implementation of a failure or error in lexical access.

As illustrated by this example, unimodal processing deficits apparent only in word comprehension can be explained quite naturally in a model relying on functional webs. A selective deficit in understanding spoken words whereas other sounds can still be interpreted, has been named *word-form deafness*. What has been said about comprehension can be extended to speech apraxia on the motor side. *Apraxia of speech* includes a deficit in producing word forms correctly. Phonological errors, omissions, hesitation phenomena, and other coordination problems in speech production predominate, whereas there is no pronounced deficit in word comprehension. Selective lesion of the motor output layer of the model network in Figure E1.2 primarily cuts neuron ensembles from their motor output fibers. This deteriorates the network's performance on the word production task, but only minimally shows in simulations of word comprehension. Similar to patients with apraxia of speech, the networks produce omissions and incorrect output in the absence of deficits in word comprehension. The simple model does not allow for modeling fine-grained aspects of apraxia of speech and word-form deafness, but it can explain the important double dissociation seen between the two types of unimodal syndromes, one specifically affecting word production and the other word comprehension.

The explanation of cases of unimodal deficits, such as apraxia of speech or word-form deafness, in which two deficits are fully dissociated (each without the other) is important. It is probably equally important to explain the fact that symptoms frequently co-occur in certain combinations. Most aphasias are multimodal deficits involving all language modalities, but each to a different degree. A simulation may therefore aim at mimicking both dissociation and co-occurrence of symptoms.

Lesion simulations using cell assembly networks suggest that the closer a lesion is to the periphery of an assembly, to the primary areas, the more asymmetrical is the disturbance. As discussed previously, a lesion in the periphery (i.e., in the input or output layer) of a network, such as the one depicted in

Figure E1.2, can lead to a unimodal deficit, either in word comphrehension or word production. However, lesion in the next layers (second from top or bottom) was found to cause a bimodal but asymmetric disturbance, for example, strong comprehension deficit but minor production impairment. Finally, a lesion in the layers in the middle of the network caused an almost symmetrical pattern of errors. The model allows for predictions on both the dissociation and co-occurrence of symptoms caused by focal cortical lesions.

When lesions were allowed in only the upper half of the network, the analog of Wernicke's area in the superior temporal lobe, the network showed strongly impaired performance on the simulated word comprehension task and relatively mild impairments when the network was used to simulate word production. In contrast, a lesion in the lower half of the network – corresponding to the inferior frontal region which includes Broca's area – led to stronger deficits in the word production task and only milder impairment in the network's word comprehension. Thus, the network simulation provides a model of a double dissociation frequently seen in patients with Broca's and Wernicke's aphasias.

Further specific features of Broca's and Wernicke's aphasias were also present in the model. For example, after lesion in the network's Broca area, activity frequently extinguished in the production task, thus simulating inability to produce a word, which is indeed seen in many patients with Broca's aphasia. In contrast, lesions on the Wernicke side of the network rarely produced such lexical omissions but rather yielded activation of not stimulated (incorrect) ensembles. This nicely corresponds to the frequent use of incorrect words by patients with Wernicke's aphasia. (For more details about these simulations, see Pulvermüller, 1992; Pulvermüller & Preissl, 1991).

These simulations explain one important double dissociation – the differential degradation of word production and comprehension in Broca's and Wernicke's aphasias on the basis of a simple model of functional webs distributed over perisylvian areas. Because individual layers of the network can be likened to areas in the cortex, the network simulation has implications regarding the location of cortical lesions that cause particular deficits.

Because a distributed model without functional segregation between modules was able to imitate a double dissociation pattern, it follows that no modular structure is necessary for explaining this neuropsychological phenomenon. This point has been made earlier (Plaut & Shallice, 1993; Pulvermüller & Preissl, 1991) and is now widely agreed on. One should, however, emphasize that these considerations by no means rule out modular models of cognition; rather, they show that different views are possible on the cortical cause of double dissociations.

The present modeling approach is an elementary one. Clearly, it is desirable to provide more detailed simulations of cognitive processes in which more and more features of the neuronal substrate are included (e.g., Sommer & Wennekers, 2000). It is desirable to develop such more detailed neuronal models of language processes in the brain (e.g., Wermter & Elshaw, 2002), not only to demonstrate their own functionality but also to model neuropsychological data and the results of metabolic and neurophysiological brain imaging studies.

As mentioned in Chapter 4, further explanations of fine-grained category-specific neuropsychological double dissociations can be made on the basis of the assumption that word webs have distinct cortical topographies. In this context, two types of explanations are relevant. One relies on the center–periphery argument made in this section, the other on different assembly topographies (see Chapter 4). The double dissociation between agrammatism (function word impairment, perisylvian lesion) and anomia (content word impairment, extra-perisylvian lesion) has been discussed in great detail elsewhere (Pulvermüller, 1995), and other word category dissociations may well be explained along the same lines. Clearly, all of these explanations are related to and rooted in cognitive models (Warrington & Shallice, 1984). Spelling them out in the language of neurons may help improve our understanding of the relevant mechanisms.

CHAPTER FIVE

Regulation, Overlap, and Web Tails

Chapter 4 offers a neurobiological perspective on word processing in the brain. The time course and topography of cortical activation during word processing, in particular during the processing of words of different categories, is discussed in some detail. The proposal is that there is a cell ensemble or functional web for each and every word, and that words with different referential meaning may have functional webs characterized by different topographies.

Looking at words more closely, more questions arise, for example, regarding complex form–meaning relationship. Two words may share their form (e.g., “plane,” meaning “aircraft” or “flat surface”), or may sound differently but have largely the same meaning (e.g., “car” and “automobile”). There are word forms that include other word forms (e.g., the letter sequence “nor,” including “no” and “or”), and there are words whose meaning includes, so to speak, the meaning of others (e.g., “animal” and “dog”). These relationships of homophony (or polysemy), homonymy (or synonymy), inclusion of word forms in other word forms, and hyperonymy (vs. hyponymy) may pose problems to an approach relying on cortical neuron webs. How might a neurobiologically realistic model of complex form-meaning relationships look? Tentative answers are discussed in the section on overlap of representations.

Other aspects largely ignored in earlier chapters relate to the information about written language and to aspects of meaning that have been characterized as emotional or affective. This information may be related to the phonological and semantic side of words through connections of the relevant word webs either to subcortical or to additional cortical neurons. A few remarks are made about the possible circuits in the section on web tails.

A concern comes from considerations on activity dynamics in the cortex. Associative learning networks run into the problem of associating everything with everything if too much learning has taken place, or if too much activity is allowed to spread. It is argued here that a regulation mechanism is needed in autoassociative models of network memory and that such a mechanism can be important for solving the mentioned putative problems regarding words that are related phonologically or semantically.

5.1 Regulation of Cortical Activity

A brain model in the tradition of Hebb's cell assembly theory runs into a number of problems. Milner (1957, 1996) discusses some of these problems in great detail. An important problem was the result of the original formulation of *Hebb's learning rule*, according to which two connected neurons that frequently fire together increase the strength of their wiring. One can call this a *coincidence rule of associative learning* because only coincident firing of two neurons is considered to have an effect on connection strength. In a network with many links between neurons, a coincidence rule can lead to ever-increasing connection strengths so that, finally, catastrophic overactivation may take place whenever the network is being stimulated.

There are, however, tricks available that allow for solving, or at least minimizing, the problem of catastrophic overactivation. These tricks seem to be applied by the real brain to allow for effective storage of memories, or *engrams*, in the cortex. Three strategies for minimizing the overactivation problem *are* introduced after this problem itself has been explained in more detail.

5.1.1 The Overactivation Problem in Auto-associative Memories

The *overactivation problem* can be illustrated using artificial associative memory networks. A fully connected autoassociative memory consists of n neurons and reciprocal connections among all of them. This network type is not a particularly realistic model of the cortex, but it can be made more realistic by omitting connections between neurons, for example by reducing the number of connections of each neuron to $\sqrt{n}$ (Palm, 1982). In this case, the network is still called autoassociative, as long as it includes loops. (A network in which different neuron populations are connected in only one direction is called a hetero-associative memory.) The idea underlying the Hebbian approach is that the neurons that show frequent coincident activity, for example, the neurons activated by an object stimulus, link into a neuronal assembly so that finally there are many object representations in the network. Each object would be realized as a neuron ensemble with particularly strong connections between its members. These ensemble–internal connections are much stronger than the average connection strength in the entire autoassociative network. Now, the problem is that coincidence learning leads to an increase in average connection strength. After much learning and strengthening of connections, so many neuronal links may have become so effective that any stimulation of some of its neurons activates the entire network, a process that may be analogous to an epileptic seizure in the real nervous system (Braitenberg, 1978a). This is undesirable because the aim is to keep representations separate and allow for retrieving, or specifically activating, individual object representations. This property is lost if too many too strong links developed.

5.1.2 Coincidence vs. Correlation

Obviously, a first strategy to avoid the overactivation problem would be to avoid the learning rule proposed by Hebb, or at least to modify it. There is good motivation for this. It has become clear that not only coincidence of neuronal firing has an effect on synaptic efficacy (Brown et al., 1996; Rauschecker & Singer, 1979; Tsumoto, 1992). Although coincident activity of two neurons can lead to synaptic strengthening, antiphasic activation of the neurons can reduce their connection strength. If one neuron fires while the other is silent, their synapse may become weaker, a process sometimes called *Anti-Hebb learning*. The correlation rule mentioned in Chapter 2 captures important features of the known principles underlying synaptic plasticity, in particular, Hebbian and Anti-Hebb learning. The correlation rule implies that each learning event includes strengthening of certain synapses (the synapses between coactivated neurons) as well as a weakening of other synapses (the synapses between activated and inactive neurons). Thus, if parameters are chosen appropriately, the average connection strength does not increase when engrams are being stored in an autoassociative network. This would make it less likely that, after substantial learning has taken place, the network shows catastrophic overactivation.

5.1.3 Sparse Coding

A second strategy for minimizing the overactivation problem is to use what has been termed *sparse coding*. This means that, among the many neurons in a network, only a small subset of neurons contributes significantly to the representation and processing of each individual engram stored in the network. The activity state of the entire network can be coded using a vector of numbers, with each number giving the activity value of one neuron of the network. For sparse coding, each engram can be characterized by only a few high values (for example, 1, meaning “strongly active”), whereas most other values are low (0, for example, meaning “inactive”). One may argue that, in this case, changing low values to high values may have the same effect in deteriorating the engram as lowering high values. However, this is not correct for an autoassociative network, in which engrams are sparsely coded. Take the example of an autoassociative network of 100 neurons in which 10 engrams are stored, each by way of 10 different neurons. The 10 neurons of each engram would be strongly associated and their common activation would be considered the retrieval of the engram. Evidently, if nine of the active neurons in the vector defining the engram were not active, the activation of the remaining 1 neuron would now not allow for engram retrieval. In contrast, if the 10 were activated and, in addition, nine randomly selected other neurons defined as inactive by the engram vector would be active – for example, as a result of noise, the engram could still be retrieved. The noise would likely activate other engrams minimally, so that they could not compete with the fully active cell assembly. Only if the nine other neurons were part of the same engram, unlikely in the case of noise, this then strongly excited ensemble could compete with the fully active engram (but would still not reach its activity level). Still, the same engram would be retrieved. Therefore, sparse coding implies that, on average, inactivating neurons defined as active by an engram vector has a stronger effect on the retrieval outcome than has the activation of neurons specified as inactive by the vector. One can therefore consider the neurons specified as active by the engram vector as primarily relevant for representing it.

These considerations underlie the postulate crucially that objects and words are represented as cell assemblies or functional webs. If words are assumed to be represented as functional webs, this has the implications (i) that each word engram is defined by an activity vector specifying the activity state of all neurons in the network (i.e., the cortex or brain), and (ii) that sparse coding is used. In other words, one cortical cell assembly is assumed to include only a small percentage of the neurons in the cortex. Estimates of actual cell ensemble size actually range between thousands to 1 million cortical neurons (Palm, 1993b). This is consistent with the sparse coding assumption.

Most cortical neurons usually fire at a low rate of a few action potentials per second. They become more strongly active only in response to specific stimuli and are not much affected by the rest (Hubel, 1995). These facts are consistent with and therefore support the notion of sparse coding of representations in the cortex. Furthermore, a welcome property of sparse coding is that the memory capacity of an autoassociative memory is largest if sparse coding with small overlap of representations is applied (Palm, 1980, 1990).

In sum, sparse coding not only has computational advantages but is also likely to be used for information storage in the cortex. One of its important consequences is the generally low activity level in the associative network, which further reduces the likelihood of overexcitation.

5.1.4 A Cybernetic Model of Feedback Regulation of Cortical Activity

Apart from correlation learning and sparse coding, a third strategy is likely to be used by the brain to prevent overexcitation. Even in an autoassociative network with moderate average connection strength, strong stimulation from outside may overactive the neurons. However, lack of stimulation may cause activity to die out. To prevent these undesired outcomes, a regulation mechanism designed to keep the activity level within certain bounds must be postulated. In the real brain, this mechanism would control the activity status of the cortex. The regulation mechanism would respond to strong activity increase by a process of inhibition or removal of background activity, that is, disfacilitation. If activity is about to die out, the control device may provide more background activity to elevate the excitation level again.

An important mechanism for controlling excitation is provided by the small inhibitory cells in the cortex itself. As mentioned in Chapter 2, local inhibitory circuits can reduce activity in small pieces of gray cortical matter if much excitation is provided by local pyramidal neurons. Thus, overexcitation in the cortex appears to be counteracted by a local feedback regulation mechanism.

A different more global control mechanism has been formulated in terms of a cybernetic feedback–regulation circuit (Braitenberg, 1978a). The feedback regulation device would take the global activity level, A, in the cortex – or any other part of the brain, B – as its input and compute an output depending on how far this measured activity value A deviates from a preset target value. The output of the regulation is fed back to the cortex. This feedback value is called θ, which is calculated, at each point in time, as a function of the activity state A. It can be conceptualized as global background activity that is increased or decreased to move A toward the target value. Braitenberg (1978a) called this mechanism *threshold control,* a term that may suggest that the parameter actually changed by the regulation device is the threshold of individual cortical neurons. However, as pointed out by Braitenberg, it is likely that membrane potentials are the variable influenced by feedback control in the brain. If the threshold of a neuron is defined in terms of the change in membrane potential needed to activate the cell, an increase in background activity lowers the threshold and a reduction in background

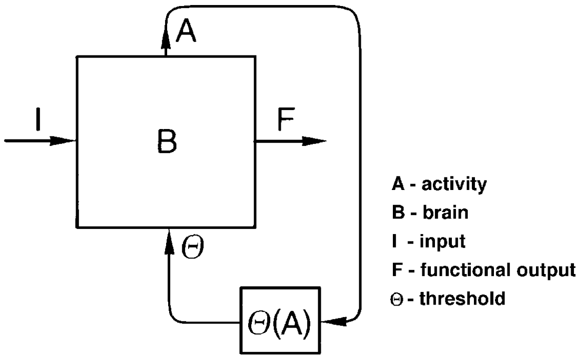

Figure 5.1. The idea of threshold regulation as formulated by Braitenberg. A brain (B) receiving input (I) and producing a functional output (F) has a second output that includes information about its activity state (A). Based on this second output, threshold values (θ) of the neurons in the brain are recalculated and fed back to the brain. Feedback control is such that higher activity in the brain raises the threshold and lower activity attenuates it. A feedback control mechanism is necessary for keeping brain activity within certain limits. Adopted from Braitenberg, V. (1978). Cell assemblies in the cerebral cortex. In R. Heim & G. Palm (Eds.), *Theoretical approaches to complex systems.* (*Lecture notes in biomathematics, vol. 21*) (pp. 171–88). Berlin: Springer.

activity elevates it again. The threshold control mechanism postulated by Braitenberg is illustrated in Figure 5.1.

Although there is agreement that regulation of cortical activity is necessary, the exact characteristics of the mechanism and the brain systems that realize it are still under discussion. For example, the threshold control mechanism regulating the activity state of an autoassociative network may take several alternative variables as its input. It could look at the actual number of neurons active in a given time window or at the average frequency of their action potentials. When it comes to specifying the working of the mechanism in more detail – as shown in Chapters 10–12 on grammar mechanisms – it is assumed that the threshold control mechanism detects fast activity increases and provides global inhibition if activity raises substantially. In particular, the control mechanism is assumed to become active if a previously inactive neuronal ensemble ignites. Furthermore, the assumption is that excitation is provided if activity levels are generally low. This is one of several realistic possibilities to model a threshold regulation mechanism.

There are alternative or complementary, proposals highlighting the possible role of cortico-cortical (Milner, 1996), cortico-striatal (Wickens, 1993), and hippocampal (Fuster, 1995; Miller, 1991) connections in controlling the activity level in the neocortex. It is not clear, however, which of these brain structures is most crucial for regulating the activity level in the neocortex. It may well be that more than one of the well-known loops formed by the cortex and other brain structures play a crucial role as regulation devices.

5.1.5 Striatal Regulation of Cortical Activity

One putative circuit is now described in more detail to illustrate how a regulation mechanism might operate. The neocortex is intimately linked to subcortical brain structures. An important loop is formed by projections form the cortex to the neostriatum (*Putamen* and *Nucleus candatus*), from there to the paleostriatum (or Pallidum), and finally to the thalamus, from where projects run back to the cortex, in particular the frontal lobes. Two of these links are inhibitory, form neostriatum to paleostriatum and from there to thalamus, so that the two inhibitory connections in series produce an activating effect. Therefore, cortical activation causes additional cortical excitation through this striatal–thalamic loop. Among the structures involved in the loop, the neostriatum is known to include many neurons that can inhibit their local neighbors and, by way of indirect connections, can have an inhibitory or excitatory effect on more distant neostriatal neurons as well. It thus appears that if neurons in the neostriatum are being activated by their cortical input, a complex pattern of selective inhibition and excitation is being produced (Wickens, 1993).

In essence, there is a subcortical loop, including neostriatum, pallidum, and thalamus, through which the cortex can stimulate itself. In addition, there is the inhibitory network in the neostriatum that can be the basis of competition. This architecture has been proposed to realize a regulation mechanism also affecting cortical activity dynamics (Wickens, 1993). One proposal was that distributed cortical neuron ensembles are linked to small populations of neurons in the neostriatum, paleostriatum, and thalamus, and that each cortical neuronal assembly self-activates through the cortico–striato–thalamic loop. Given that each cortical assembly has this subcortical extension, the striatal inhibitory neurons can provide the basis for competition between the different ensembles. This is schematically illustrated in Figure 5.2, in which open and filled circles are thought to represent two neuron ensembles. Note that the two overlap in their cortical part, but inhibit each other by way of the inhibitory connections between their neurons in the neostriatum (Miller & Wickens, 1991; Wickens, 1993).

These considerations demonstrate that feedback regulation of cortical activity can be implemented in the brain. Proposals for such implementation are available and in agreement with cortical structure and function. The exact wiring must be clarified by future research.

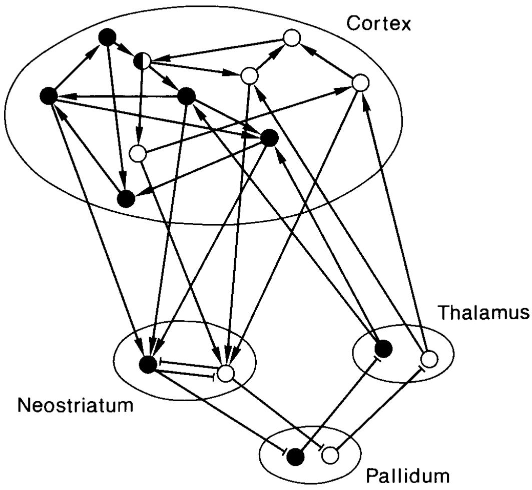

Figure 5.2. A possible realization of a regulation device in the brain is illustrated. Cortical neuron ensembles are assumed to be strongly connected to subcortical neurons in the neostiatum, paleostriatum, and thalamus. Each cortical neuronal assembly can activate itself through this subcortical loop. Inhibition between cortical neuronal assemblies is provided indirectly by the inhibitory connections in the neostriatum. Filled and open circles indicate two overlapping webs. The subcortical inhibition can prevent simultaneous ignition of both overlapping webs.

5.1.6 Summary

The problem of cortical activity regulation can be minimized if correlation learning, sparse coding, and feedback regulation are being applied. This seems plausible neurobiologically. The basic idea therefore becomes that correlation learning leads to the formation of sparsely coded representations of objects and words, each by one functional web, and that feedback regulation keeps the cortical activity level within certain bounds.

The feedback regulation mechanism can be described in cybernetic terms, and there are also proposals to spell it out in the language of neuronal circuits. One possibly is that the inhibitory neurons in the neostriatum play a role. Full activation of one of the neuron ensembles would, by way of striatal inhibition, prevent other competing networks becoming active as well. In case of lack of strong activity in the cortex, the positive feedback loop through subcortical nuclei could enhance the cortical activity level gradually. The inhibition between neuronal assemblies can be provided even if the cortical ensembles overlap – that is, if they share neurons. The illustrated circuit shows that it is possible to keep overlapping cortical representations and activity patterns functionally separate. It is therefore possible to realize functionally discrete engrams by overlapping neuron ensembles.

5.2 Overlapping Representations

How would functional webs representing related words be organized? A similar question is, how can different readings of the same word or two different ways to realize it phonologically, be modeled? The general answer proposed for these questions will be that related items are organized as overlapping cortical neuron ensembles. Putative mechanisms for such overlapping representations have been specified in Section 5.1. In general, these mechanisms allow each neuronal representation α to become fully active, whereas, at the same time, preventing those neurons of other overlapping ensembles β, $\gamma, \ldots, \omega$ that are not included in α from becoming strongly active as well. In the proposed jargon of functional webs, the more precise formulation is the following. The ignition of α prevents other webs $\beta, \gamma, \ldots, \omega$, some of which overlap with α, from igniting at the same time. The overlap areas α shares with $\beta, \gamma, \ldots,$ or ω would be activated as part of the ignition of α.

The following proposal is that overlapping but mutually exclusive distributed representations underlie the processing and representation of different types of related words. As pointed out in Chapter 4, functional webs representing individual words may include

- a phonological or word-form related part located in perisylvian language areas and strongly lateralized to the left
- a semantically related part distributed over various areas of both hemispheres, whose topography may depend on semantic word properties.

The two parts, the form/phonological and semantic subwebs, would be functionally linked, and thus exhibit (i) similar dynamics and (ii) mutual functional dependence.

In this framework, the representation of phonologically and semantically related words is straightforward. Semantically related words should overlap in the semantic, widely distributed, bihemispheric part of their functional webs, whereas phonologically related words should share neurons in the perisylvian lateralized part of their functional webs.

The overlap between representations also assumed by many other psycholinguistic theories (Section 6.2) explains why words closely related in meaning and form can influence or prime each other. In this context, to *prime* means that, in experiments investigating word processing, the presentation of one of the related items, also called the *prime*, influences the processing of the other related item that is presented later, usually called the *target*. Priming effects can be facilitatory or inhibitory, that is, the prime word can improve or deteriorate the processing of the target word. Both semantically and form-related words can have a facilitatory priming effect on each other (Humphreys, Evett, & Taylor, 1982; Meyer & Schvaneveldt, 1971). It must be mentioned, however, that the paradigm and, in particular, the onset with which the stimulus words are being presented can have an effect on whether a priming effect is obtained, and on whether it is facilitatory or inhibitory (Glaser & Düngelhoff, 1984; Levelt et al., 1991; Mohr & Pulvermüller, 2002).

The issues of semantic and form-related overlap of functional webs is discussed in greater depth in the following subsections.

5.2.1 Homophones and Form-Related Words

Syllables such as "spy" have at least two possible meanings with only minor, if any, relation to each other. How would such homophonous (or polysemic) words be organized in a model of functional webs? The syllable may occur frequently in the contexts of spiders and secret agents. The web representing the word form could, therefore, develop connections to the neuronal counterparts of both word meanings. As argued in Chapter 4, these word-form related or phonological webs should be localized in the perisylvian cortex, whereas the meaning-related neurons would primarily be expected outside the perisylvian cortex, in extra-perisylvian space. In addition, only the form-related assemblies would exhibit strong laterality.

Now the problem arises that perception of the syllable "spy" will activate both semantic representations at a time. This is not desired; the two possible word meanings should be kept separate. Conceptually, it is best to speak about two word representations, each with its specific meaning, that share their phonological representation. Likewise, it appears appropriate to postulate two neuron ensembles, one for each word, that overlap in their perisylvian phonological part.

The mutual exclusion of the two overlapping functional webs could be provided by a regulation mechanism of the type discussed in Section 5.1. A scenario for this would be as follows: A word stimulus would activate the phonological part shared by the two functional webs representing the two words. Because the semantic parts of both word webs are strongly connected to the shared phonological representation, they both receive excitation. There is a race between the two overlapping word webs that finally ends in the ignition of one of them, the one reaching its activation threshold first. The outcome of the race is likely determined by the internal connection strength of each of the overlapping webs – the web of the more frequent homophone being more likely to ignite first – and by activity in the webs due to input in the past, for example, preactivation of the web of the word that best fits into the context. The ignition of one of the word webs activates the regulation mechanism, thus preventing the ignition of the competitor. Nevertheless, as detailed, the competitor web initially profits from the activation process because it is partially activated in the initial phase of the activation process, before the ignition takes place.

This scenario is reminiscent of psycholinguistic ideas developed on the basis of priming studies (Swinney et al., 1979). Reaction time studies indicated that presentation of a homophonous word form first leads to partial activation of all of its homophone meanings, whereas only at a later point in time is one of them being selected while the others are suppressed. Strongly connected overlapping word webs that activate each other, but cannot ignite at the same time as a result of regulation processes, may be the mechanism underlying this; initial broad partial activation of all overlapping webs and later selective ignition of one of them would be the tentative neurobiological realization.

This mechanism is not necessarily restricted to real homophones or polysemes. The same process could underlie the decision between different "submeanings" or uses of one particular word. *School*, for example, can either refer to a place ("the school in Brentwood"), an event in such a place ("school begins at nine"), a body of persons ("the school of Tesnièrian linguists"), or an animal group ("a school of fish"). These readings are related, but nevertheless exclude each other in a particular context. Their putative neuronal counterparts are overlapping functional webs that share some of their neurons in perisylvian areas, but inhibit each other by way of a regulation mechanism.

Homophones share their word form. There are also words composed of other words, and similar arguments can be made with regard to them. In this case, the form of one word includes the form of another one. This is called *form inclusion* and the included word form is distinguished from the including word form. This relation of form inclusion may hold for the words' phonology, orthography, or both. Examples of word pairs one of which includes the other are, for example, *scan* vs. *scant*, *pair* vs. *repair*, and *depart* vs. *department*. In these cases, the relationship between the words concerns their formal aspects, but not their meaning. The representation of the form of the smaller item that is part of the larger word can be envisaged to be organized as sets of neurons, one of which includes the other. This idea is sketched in Figure 5.3 in the diagram on the lower left.

It may not be appropriate to postulate that the form-related neuronal populations fully include each other, because the pronunciation of the included item is usually affected by the context of the larger one. Therefore, the included item pronounced in other contexts can exhibit acoustic

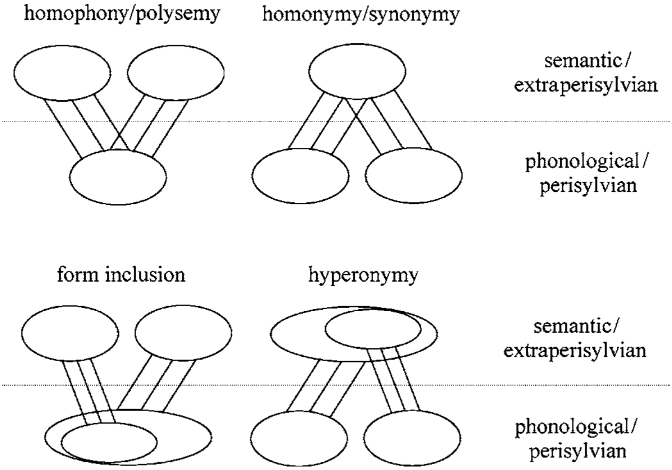

Figure 5.3. Putative organization and cortical distribution of functional webs representing related words. Each functional web is represented schematically as two ovals: one for its semantic part and the other for the phonological or word-form related part. The ovals would therefore represent sets of phonological and semantic neurons. Overlap between ovals means that webs share neurons. As discussed in Chapter 4, the semantic subassembly would be expected to have most of its neurons located outside the perisylvian areas, whereas the phonological subassembly would be placed in the perisylvian areas. Lines represent connection bundles linking neurons within a functional web. (Upper left) The diagram sketches the proposed organization of functional webs realizing *homophonous* words ("spy"), or different readings of the same word ("school"). (Upper right) Putative correlate of synonymous – or near synonymous – words ("car" and "automobile"). (Lower left) If one word form includes another word form ("repair" and "pair"; this is dubbed *form inclusion*), the relation of inclusion may also hold for the respective phonological webs. (Lower right) In the case of *hyperonymy*, if one term is more general than another (*animal* vs. *dog*), the relation of inclusion may hold for the semantic parts of the webs.

features it lacks in the context of the larger word that includes it. It may thus appear more appropriate to postulate overlapping phonological representations rather than that one is fully included in the other (as the written representation would suggest). Still, however, there would not be too many specific form features – and therefore neurons – of the small items, but several specific features – and neuronal correlates thereof – of the larger item.

It is particularly difficult for an autoassociative memory to keep separate two representations, one of which fully includes the other. A mechanism for this could be envisaged, for example, on the basis of the striatal inhibition model, as sketched in Section 5.1. However, it is certainly easier to separate two distributed representations if each includes a substantial number of neurons that distinguishes it from the other representation. Ideally, the number of neurons shared between two ensembles should be small, certainly not larger than the number of the neurons distinguishing one web from the other. The larger the percentage of overlapping neurons becomes, the more difficult it becomes to keep two representations apart.

It would appear that the brain uses an appropriate strategy for overcoming the overlap problem. Word forms that include others can be attached to different semantic representations. These semantic representations differentiating between the similar forms can, so to speak, be used as handles to keep the word forms separate. The non-overlapping parts, rather than separate neurons representing lexical items, can then be the target on which a regulation mechanism can operate in order to exclude full activation of one of the ensembles while the other one ignites. It is clear that various excitatory and inhibitory processes could occur between such partially overlapping networks of neurons if they are stimulated while activity is controlled by a regulation mechanism. Nevertheless, the overlap would make facilitatory effects possible.

Again, mutual exclusion of ignitions of overlapping neuronal representations in a regulated system profits from relatively small overlap regions. Such mutual exclusion is difficult in a system of fully inclusive neuronal ensembles. If form representations strongly overlap, it is possible to maximize the distinguishing neurons by adding completely different semantic web parts to the similar form representations. This may be so if word forms have very different meanings. Many lexicalized or semantically opaque words include at least one part of their form that can also be used in a completely different meaning. The meaning change results, for example, if a derivational affix is attached to a noun or verb stem, as, for example, in the previously mentioned pairs *repair* vs. *pair* and *department* vs. *depart*. In a model of functional webs, they would be realized as form-inclusive ensembles, as sketched in the diagram on the lower left in Figure 5.3.

The representation is different for very similar forms whose meaning is closely related and can be deduced from the meanings of the morphemes they are composed of. These semantically transparent words, such as *unpair* or *departure*, can be conceptualized as being composed of their parts at both the word form and the meaning level. The words *depart* and *departure* have the same possible referent actions. Representations of words such as *departure* would be composed of the two distinct representations of *depart* and the derivational affix (Marslen-Wilson, Tyler, Waksler, & Older, 1994). This proposal is consistent with the priming effects observed between semantically transparent composites and their submorphemes (*departure* and *depart*), which appears to be more solid and stable than priming between semantically opaque forms and their elements (*department* vs. *depart*) (Marslen-Wilson et al., 1994; Rastle et al., 2000).

Translated into the language of functional webs, this may mean that a semantically transparent complex form is organized as a widely distributed web (realizing a content word) and a more focused web (realizing the derivational affix). These two networks would be connected serially, by what is called a *sequencing unit*, or sequence detector in Chapters 9–12. Facilitation effects would be the result of the activation of the same networks by prime and target stimuli. Instead, the relation between a lexicalized or semantically opaque form and its parts is more complex, as suggested by Figure 5.3, lower left diagram. Some facilitation could, in this case, be the result of the overlap of form representations, but the competing meaning representations would work against it.

5.2.2 Synonyms

The mirror image, so to speak, of homophony is homonymy or synonymy. Two different word forms share their meaning – or are very similar in meaning. Examples would be *automobile* and *car*. The relation between synonyms may be realized cortically by functional webs sharing, or largely overlapping in, their semantic, mainly extraperisylvian, part.

A decision between the two overlapping representations is necessary when the semantic part shared by the neuron ensembles are activated. If a picture must be named, the overlap region of two synonyms (or near synonyms) would be activated to the extent that some activity would also spread to each of the word-form representations. The best activated word web would ignite first, thus inhibiting the competitor through the regulation mechanism. Again, activity already present in the webs at the start and their internal wiring strength may be crucial for deciding which of the alternatives is being selected. The average internal connection strength of the web is likely influenced by the frequency with which the word is used in language.

The upper right diagram in Figure 5.3 gives an idea of how the representation of synonyms may be realized in the brain; by identical or strongly overlapping widely distributed representations of the meaning and distinct perisylvian left-lateralized subwebs for the different word forms. This is essentially the mirror image of the homophony/polysemy diagram on the lower left. Therefore, one may propose that the general activity dynamics and interference effects should resemble each other. This is, however, not necessarily the case because the overlap of homophone representations is necessarily with respect to a temporally structured neuronal unit, whereas a possible temporal structure in a semantic web part may be disputed (see Chapter 8).

A word can be realized in different ways. Different realizations of the same word, for example, its pronunciations in different dialects, can be conceptualized using a similar mechanism as proposed for synonyms. In contrast to synonyms, a strong overlap should exist not only in the semantic parts of the webs, but also in the phonological parts. Chapter 8 addresses the issues of alternative realizations of the same word form.

The meaning of words can be organized hierarchically. Words such as *animal*, *dog*, and *greyhound* refer to wide and more narrow categories of objects, and therefore a model based on functional webs would conceptualize their meanings as neuron sets that include each other. The larger category – to which the *hyperonym* ("animal") refers – would include a smaller category – referred to by the *hyponym* ("dog"). The semantic part of the web of the hyperonym would therefore include the semantic part of the hyponym's web. This is illustrated by the diagram on the lower left of Figure 5.3. Clearly, the overlap of semantic representations should allow for facilitatory priming effects, whereas the alternative word forms can be the basis of competition.

5.2.3 Prototypes and Family Resemblance

The claim that a more general term's semantic web is the union of all semantic representations of its hyponyms is in need of refinement. The lower-level categories share features, and the respective semantic representations can therefore be conceptualized as sets of semantic feature neurons that overlap. The name of the higher-level category can be used to refer to all objects the hyponyms can refer to. Therefore, the higher-level category name correlates best with this overlap of the hyponyms' semantic features. Given correlation learning is crucial, the connections between the word form of the more general term and its semantics relies primarily on the strong connections between word form web and the semantic overlap neurons.

Figure 5.4 schematically represents the shared semantic features by the intersection of three ovals. The visual features of the mouths or eyes of typical frequently encountered animals, such as dog, cat, and mouse, may be thought to be located here. If a less prototypical animal lacks some of these features, or even most of them – think of the words *octopus* or *jellyfish* as examples – its semantic representation would not link into this overlap region (dashed oval in Fig. 5.4). It is therefore clear that priming between the hyperonym and the less prototypical hyponym would be much reduced compared to the priming expected between hyperonym and prototypical hyponyms.

If a picture of an object is being presented to a person and he or she is asked to produce a name for it, the activation race between possible word representations would again be determined by their internal connection strengths. The outcome of the race depends on the correlation among the relevant neurons and, thus, for example, on the frequency with which the

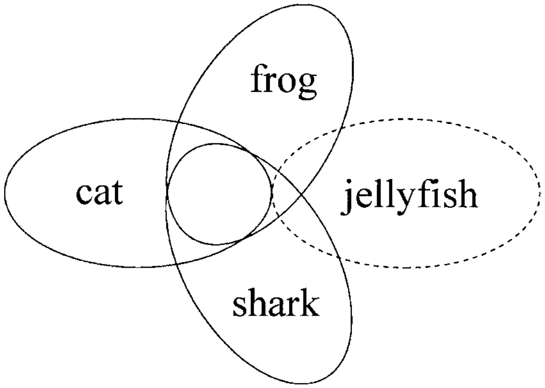

Figure 5.4. Putative organization of semantic relationships between concepts: overlap of prototypes and family ressemblance. Ovals represent semantic web parts of individual words. *Overlap* means shared semantic features between word representations and shared feature neurons between webs. Think of the overlap area to include neurons responding to visual features of mouth, eyes, and heads of typical animals. Three overlapping semantic web parts of prototypical concepts are illustrated by ovals in solid lines (e.g., *cat*, *frog* and *shark*). A fourth semantic web representing a nonprototypical concept may not link into the overlap area of the three prototypical webs, but nevertheless exhibits family ressembence with them – that is, it would share some semantic feature neurons with some of the other webs (e.g., *jellyfish*). The union of all of these semantic representations may be considered the semantic part of a web realizing a hyperonym of the words whose meanings are represented by the individual ovals (e.g., *animal*). However, the sound-to-meaning correspondence of the hyperonym would be realized primarily as connections between the overlap area and its word-form representation.

respective words are used (their *lexical frequency*). Frequently occurring basic-level category names, such as *dog*, may therefore be preferred to rare, more specific terms, such as *greyhound*. A higher-level category name will lose the race against a more specific term (*animal* vs. *dog*) primarily because of the better correlation between the specific set of activated semantically related features on the one hand and the basic category name on the other. An additional factor is that higher-level categories can refer to very different objects or actions, as Wittgenstein (1953) illustrated using the word *game* as an example. The correlation among the semantic features these words relate to is therefore necessarily low, resulting in relatively weak average links within the semantic part of their webs. The word web of *dog* would therefore ignite faster than that of the more general term *animal* – because of its better meaning-form correlation and also faster than the web of the more specific term *greyhound*, because of its higher lexical frequency. (Additional factors may also be relevant.) These considerations are reminiscent of the priority of basic level categories of the *dog* type in cognitive processing as made evident by behavioral testing (Rosch et al., 1976).

The illustration of overlapping meaning representations in Figure 5.4 has further implications for a model of functional webs or other types of strongly connected neuron ensembles. It shows that the borders of a functional web are not necessarily easy to define (formal definitions of *neural assembly* have been discussed by Palm, 1981). The word *animal*, whose meaning may be thought of as the union of all the features of its hyponyms, would have the union of the semantic web parts of its hyponyms as its neuronal equivalent. In contrast to this wide definition, one may argue that only the overlap region of semantic neurons shared by all hyponym webs should represent the "core meaning" of the more general term. The problem with this latter view is that the overlap of all hyponyms' semantic webs may be zero, or at least not very specific.

This is essentially Wittgenstein's (1953) argument about family resemblance. The meanings of a word in different contexts do not necessarily have a common kernel of semantic characteristics (or criteria). The different meanings may just resemble each other like members of a family, without sharing a set of features or genes. Likewise, the semantic neuron sets representing the different context-dependent meanings of the word *animal* may not have a significant overlap. The situation may just be as illustrated in Figure 5.4, in which the uncommon meaning (e.g., if the word *animal* is used to refer to jellyfish) is linked into the other semantic web parts but not their intersection.

This raises deeper questions about the definition of cell assemblies and functional webs. The implication of these considerations are that a cell assembly or functional web should not be conceptualized as a static neuron set with sharp boundaries (Braitenberg, 1978a; Posner & DiGirolamo, 1999). Depending on the general background of cortical activity and previous activation of the web, more or less neurons may become involved in an ignition of a given functional web.

These considerations show that the view "functional web = word" raises theoretically interesting questions. A necessary conclusion is that functional webs should not be thought of as constant neuron sets with sharp boundaries that always ignite and reverberate in the same way. Although a core of word form and semantically related neurons may always be included in the ignition, there are neurons at the border of the set, so to speak, that may or may not take part in an ignition. Their inclusion into or exclusion from the ignition process would depend on context – that is, on background activity and activity already present in the web. As an example, the core of the semantic ensemble could be defined as the neuron set shared by at least two of the four overlapping ovals in Figure 5.4, representing the more typical features of animals. A similar point is made with regard to phonological features in Chapter 8. Chapters 10–12 focus on words used in different grammatical contexts and how they may be modeled in a network composed of functional webs.

5.3 Web Tails

This section discusses two very different issues that may nevertheless have a similar neurobiological basis: How may the cortical representation of written words link into the functional webs set up for words and what may be the putative neurobiological correlates of affective meaning?

5.3.1 Affective and Emotional Meaning

The idea that cortical neuron webs may be linked to additional subcortical circuits is discussed in Section 5.1. Evidence for the involvement of thalamic and striatal circuits in language processing comes from clinical studies. Further, emotional and affective word properties may be represented by neurons in the amygdala and midbrain. As a result of correlation learning, these subcortical neurons may be attached as a tail, so to speak, to the cortical neuron ensembles representing words (Pulvermüller & Schumann, 1994). There have been several tentative explanations of the variable success of learners who started to learn their second language late in life. One proposal assumes differential attachment of amygdala–midbrain tails to word webs. Only if word processing and subcortical emotionally related brain activity correlate, the learners' word webs were envisaged to grow amygdala–midbrain tails. Activation of these networks would flood the forebrain with dopamine, thereby possibly facilitating further language-related learning. Therefore, a person learning a second language who has already acquired a few word webs with amygdala–midbrain tails may facilitate storage of more information, including language and grammatically related information, when these networks are active.

Although the proposal has led to research efforts that yielded data consistent with the basic idea at both the behavioral level (Schumann, 1997) and that of neurophysiological investigation (Montoya et al., 1996), a systematic investigation using functional brain imaging techniques must await future research.

5.3.2 Linking Phonological, Orthographical, and Meaning-Related Information

In Chapters 4 and 5, the visual representations of written words were largely ignored. The webs realizing word forms were proposed to be phonologically related. Clearly, however, literate speakers must have acquired learned representations of written words and their neurobiological counterparts are likely localized in visual areas of the occipital and inferior temporal lobes. Equally clearly, however, the written word-form representations should be

Ellis and Young, 1988 Connections proposed From ↓ to →	Auditory analysis level	Phoneme output level	Auditory input lexicon	Speech output lexicon	Semantic system	Visual input lexicon	Graphemic output lexicon	Visual analysis system	Grapheme output level
Auditory analysis level	●	❍	❍						
Phoneme output level	❍	●		❍					❍
Auditory input lexicon	❍		●	❍	❍				
Speech output lexicon		❍		●			❍		
Semantic system			❍	❍	●	❍	❍		
Visual input lexicon				❍	❍	●		❍	
Graphemic output lexicon							●		❍
Visual analysis system		❍				❍		●	❍
Grapheme output level									●

Autoassociative Network model Connections proposed From neurons representing ↓ To neurons representing →	Acoustic features of sounds-in-context	Articulatory feature of sounds-in-context	Acoustic word form	Articulatory word form	Word meaning	Visual word form	Writing pattern of word form	Visual features of letters-in-context	Features of writing gestures-in-context
Acoustic features of sounds-in-context	●	❍	●	❍	❍	❍	❍	❍	❍
Articulatory feature of sounds-in-context	❍	●	❍	●	❍	❍	❍	❍	❍
Acoustic word form	●	❍	●	❍	❍	❍	❍	❍	❍
Articulatory word form	❍	●	❍	●	❍	❍	❍	❍	❍
Word meaning	❍	❍	❍	❍	●	❍	❍	❍	❍
Visual word form	❍	❍	❍	❍	❍	●	❍	●	❍
Writing pattern of word form	❍	❍	❍	❍	❍	❍	●	❍	●
Visual features of letters-in-context	❍	❍	❍	❍	❍	●	❍	●	❍
Features of writing gestures-in-context	❍	❍	❍	❍	❍	❍	●	❍	●

coupled to the perisylvian phonological webs, and this should be so for all word categories alike. Therefore, no predictions on word-category differences can be derived, except if the correspondence of written and spoken word form differs, as is the case for words with regular and irregular spelling (*home* vs. *come*).

Correlations are present at the levels of features of letters and phonemes and at the level of spoken and written word forms. Obviously, letter–sound correlation plays a greater role for regularly spelled words than for irregular ones. Additional correlation exists between aspects of the written word form and aspects of referential meaning. The correlation between individual letters and meaning aspects might not be high in most cases; however, if letters and their local context are assumed to be processed as units below the word and morpheme level, relevant correlation between letters-in-context and meaning aspects can also be postulated. The proposal is that neurons related to context-sensitive variants of a phoneme or grapheme would link with semantic neurons. This suggestion is worked out in greater detail with regard to phonemes in Chapter 8.

Neuropsychological models in the tradition of Morton's logogen model (1969) have shed light on the complex relationship between systems involved in the processing of input- and output-related information about spoken and written words. *Autonomous processors*, modules for the production and perception of spoken and written words, were postulated and related to modality-specific lexical modules and one central semantic system (Ellis & Young, 1988).

The assumption of a uniform central semantic system contrasts with other approaches in which much emphasis is put on multiple semantic systems processing different types of category-specific semantic information. Chapter 4 summarized such an approach.

Figure 5.5. (Upper matrix) Matrix display of the core part of the modular model of word processing proposed by Ellis and Young (1988). The terms in the rows and columns refer to individual modules. Each dot represents a connection from a module indicated in the left column to a module indicated in the top line. Filled dots refer to within-module connections. Open dots refer to connections between modules. Only 29% of the possible between-module connections are realized. (Lower matrix) If information about spoken and written word forms and about meaning is thought to be processed by neurons in an autoassociative memory, all possible connections should be present and effective. In this diagram, proposed local cortical connections are indicated by filled dots and long-distance cortical connections are indicated by open dots. The connections can be taken to define the internal structure of one word web. The matrix lacks detail about important aspects of the neurobiological model. For example, it does not specify category-specific semantic systems and glosses over the fact that phonologically related neurons may be included in (rather than connected to) the web realizing the word form. Nonetheless, the matrices stress a potentially important difference between a modular approach and an autoassociative memory model.

The main proposal of modular models of word processing is that the mentioned types of word-related information are processed in separate modules with specific selective connections between them. Figure 5.5 illustrates the modules and connections postulated by one of the most developed modular models of word processing (Ellis & Young, 1988). The model is presented here in the form of an autoassociation matrix in which dot signs represent connections and blank squares represent their absence. Each term appearing on the left (and again appearing at the top) refers to a distinct functionally autonomous system, a *module*. It can be seen that more than half of the 81 possible connections between modules are assumed to be not present. It appears appropriate to assume very strong connections within each of the nine modules, because information processing is primarily taking place there. These within-module connections are indicated by the black dots on the diagonal. Between modules, only a minority of the possible connection is proposed to be realized (21 out of 72 possible ones, which equals 29 percent). The sparseness of connections is motivated by the research strategy applied. The zero assumption was no connection, and each connection was required to be motivated by a double dissociation or by at least one patient whose behavior was interpretable as evidence that the respective type of information transfer was specifically impaired. Thus, insertion of an open dot (or arrow in the original diagrams) needed justification by neuropsychological dissociations.

An associative memory model of the respective types of information exchange would suggest a different conceptual approach. The cortex would be considered an autoassociative memory in which correlation causes strong links. It cannot be disputed that writing and seeing a letter is accompanied by neuronal activity in visual and motor cortical areas, including primary but probably extending into higher-order areas. Because as a rule the frontal and temporo-occipital systems are strongly linked and connections are reciprocal, it makes sense to postulate that strong reciprocal connections develop between the visual and graphic representations. This leads to the proposal that additional reciprocal connections between writing gestures and visual analysis (not present in the original matrix) should exist. Furthermore, young children learning to write would usually articulate and/or hear the respective phonemes/graphemes. Thus, reciprocal connections between auditory and visual phoneme and grapheme systems should develop as well. This leads to the addition of four more connections, linking the two analysis systems to each other and to the output systems also. The same argument should hold at the level of word forms, and the semantic system should have reciprocal connections to all other systems. A difference between models would be that the cotext-sensitive variants of phonemes, and possibly also graphemes, would be thought to be part of (rather than connected to) the word-form representations in the respective lexicons.

In summary, the postulate put forward here is that strong connections are present between all components of a distributed representation of a word. The zero assumption would therefore be that neurons involved in the processing of one aspect of a word are linked directly to neurons involved in the processing of all its other aspects. Neuropsychological dissociations can still be explained within the autoassociative approach, because cutting one particular connection – together with the partial lesion of adjacent representations – may still cause a processing deficit primarily affecting performance on specific tasks (see also Excursus I).

In conclusion, a neurobiological model can incorporate aspects of visual word processing. The neuronal webs may include attachments or tails that are connected reciprocally to the neuron sets representing the phonological word form and the word's semantics. It must be left for future research to show how an autoassociative network model can account for the great variety of neuropsychological symptom constellations involving reading and writing.

5.4 Summary

This chapter aimed at refining aspects of the functional web model of language processing outlined in Chapter 4. Mechanisms for maintaining the cortical equilibrium of activity were discussed in Section 5.1, and complex sound-to-meaning relationships of words were treated in Section 5.2. Finally, the possibility was mentioned that word webs may have additional neuron populations attached to them that could relate to knowledge about the affective valence of the words and to the written image of a word and respective actions involved in writing it.

The regulation mechanism discussed in Section 5.1 plays an important role in the grammar mechanisms discussed in Chapters 10 to 12. The regulation mechanism establishing inhibition between strongly connected neuronal ensembles can be considered a mechanism that explains a universal feature of language occuring at various levels. It could be important for the mutual exclusion of two word forms with the same meaning, of two pronunciations of the same word, of two meanings of homophonous words, of two readings of the same word, and even of two interpretations of the same sentence (*word string*).

CHAPTER SIX

Neural Algorithms and Neural Networks

What does the investigation of artificial models of networks of neurons contribute to the investigation of brain function in general, and of language mechanisms in particular? There are at least two possible answers to this question.

One answer is that an artificial neural model can be used to prove that a circuit of a certain type can solve a given problem. When thinking about complex interactions of functional neuronal units, one is in danger of losing track of what a network actually can, and cannot, do. Here, a simulation may help by providing an existence proof that a desired result can be obtained using a given circuitry or circuit structure. A successful simulation never proves that the network it is based on is actually realized in the nervous system; it shows only that the network type the simulation is based on is one candidate. To call it a realistic simulation, other criteria must be met: in particular, that the network structure can be likened to brain structure and that the functional principles governing the neurons' behavior and their actual behavior have analogs in reality as well.

A second possible answer to the question about the significance of neuron models is that they can serve as illustrations of one's ideas about brain function. In the same way as a detailed verbal description or a description in terms of algorithms, a neuron circuit can help to make an idea more plastic. However, algorithms or verbal descriptions of mechanisms cannot always be likened to processing elements in the brain. In contrast, model circuits always have some relationship to potentially real neuronal mechanisms, and it is possible to make such a model more realistic by including more detail about neuroanatomy and neurophysiology into the network's structure and functioning.

Historically, neural networks have been of utmost importance in a discipline one may want to call *brain theory*, the systematic theoretical investigation of the function of the brain. Groundbreaking work started in the 1940s with McCulloch and Pitts' work, in which symbols called *artificial neurons* were used to illustrate ideas about how neurons may realize logical operations. This chapter introduces these ideas, gives a brief overview of some neural network architectures, and highlights a few applications to language-related problems.

Some of the network architectures featured in this chapter exhibit family resemblance to the cell assembly framework outlined in earlier chapters. One obvious difference between the neural network approaches to language discussed later and the cell assembly framework is that the latter provides explicit statements about the brain areas in which neuronal processes related to cognition are postulated. Such statements are sometimes lacking in neural networks frameworks in which cognitive processes are assigned to artificial neurons ordered in layers or modules but not to brain areas. It is, however, not a principle problem to relate neuron layers to brain structures. Differences and common features between different types of neural network proposals are discussed in more detail.

6.1 McCulloch and Pitts's Logical Calculus as a Starting Point

In the early 1940s, McCulloch and Pitts published a paper titled "A Logical Calculus of Ideas Immanent in Nervous Activity" (1943). In this article, the authors discuss putative brain mechanisms possibly underlying the ability to recognize complex sequences of events. Their "calculus" was later reformulated and further developed as a theory of finite-state automata (Kleene, 1956; Minsky, 1972). Considerations were based on circuits made up of simple computational elements that share properties with nerve cells. The question was to which type of events such artificial neurons could respond to specifically. In this context, the term *event* refers either to stimuli and stimulus constellations in the environment or to neural events – that is, activity patterns of other artificial neurons spread out in space and time. The computational units would therefore specifically respond to, or "recognize," spatiotemporal patterns of activity (and stimuli), and analogous circuits can be constructed as producing, or "generating," similarly complex events.

The proposed artificial computational elements were later called *McCulloch–Pitts-neurons.* They transform an input into an output. The output is binary, and the neuron is either active or inactive. This binary character is inspired by the knowledge about real neurons that, at each point in time, either generate an action potential as an output or not. Similar to their real counterparts, the artificial neurons are active at a certain point in time if they have received enough excitatory input one moment, or *time step*, before. For

Table 6.1. The logical operations carried out by the circuits in Figure. 6.1 are OR, AND, NOT, and EITHER–OR (or XOR). The truth conditions of these operations are summarized. Variables a and b can be thought of as statements that can be either true (T) or false (F). The table lists the resulting truth values of statements that include the operators. If T is replaced by 1 and F by 0, and a and b by the neuron labels α and β, the activity value of the rightmost "cardinal" cell in the four diagrams of Fig. 6.1 can be derived from the activity states of the leftmost "input neurons."

a	b	a OR b	a AND b	NOT a	a XOR b
T	T	T	T	F	F
T	F	T	F	F	T
F	T	T	F	T	T
F	F	F	F	T	F

ease of illustration, the time continuum is, so to speak, sliced into discrete time steps.

The artificial neurons vary with regard to the number of simultaneous inputs they need for becoming active. One neuron may be activated by only one input; it would have a threshold of 1, whereas others may have a higher threshold, for instance, of 2, 3, or 10, and would therefore need at least 2, 3, or 10 simultaneous excitatory inputs to become active. The artificial neurons are usually symbolized by circles, which can be thought to represent the analog of cell body plus dendrites of real neurons, and by arrows pointing to other neurons that may be considered to be analogs of axons and their branches.

The neuron circuits one can build using these artificial threshold neurons can realize logical operations (see Table 6.1). Two neurons α and β may project to a third neuron γ, whose threshold is 1 (Fig. 6.1, uppermost diagram). This circuit can be considered a representation of a logical "Or" operation. The third neuron, γ, becomes active under three conditions: if α is active, if β is active, or if both α and β are active at the same time step. If any of this happens at time t, the neuron γ fires at the next time step, $t+1$. Otherwise it remains silent, given it does not receive additional input from other neurons. Activity of the neuron γ would therefore signify the event that α or β were active just one time step earlier. One can say that the circuit realizes, and the neuron γ represents, the logical "Or" operation.

The neurons α or β may be sensory neurons that become active when specific stimuli are present in the environment. The stimuli, or stimulus features, may be labeled a and b, respectively. In this case, the activity of the γ neuron

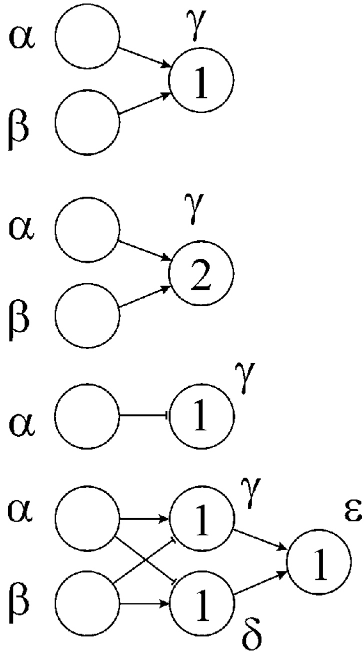

Figure 6.1. Logical circuits. In each diagram, the right-most cardinal neuron specifically responds to an activity constellation of the input neurons on the left. The cardinal neuron becomes active if neuron α OR neuron β is active (uppermost circuit), if neurons α AND β are simultaneously active (second from top), if neuron α was NOT active (third from top), and if EITHER neuron α OR neuron β have been active ("XOR" circuit at the bottom). Circles denote artificial neurons and arrows directed connections of strength 1. The activation threshold is indicated by a number in the circle. (For further explanation, see text.)

would represent the complex event that *a* or *b* was present in the input. The event *c* indicated by the γ neuron's activity would be equal to "*a* or *b*." To choose a more concrete example, the sensory neurons may respond to stimuli appearing at the same location of the visual field. One of them could be responsive to green color and the other one to circle shapes. In this case, the firing of the γ neuron would tell one that an object with the feature of greenness or roundness was present. Therefore, it would fire if the object was a green ball, an orange, or a crocodile.

By changing the threshold of the γ neuron to 2, the function of the circuit can be altered (Fig. 6.1, second diagram from top). In this case, two simultaneous inputs, from neurons α and β, respectively, are necessary to activate γ. This neuron would now only respond after both sensory neurons had been active at the same time. Thus, the stimulus features *a* and *b* must both be present. The neuron γ would respond in the event of the perception of a green ball, but not to an orange or crocodile. This circuit now realizes a logical "And" operation.

The logical "Not" operation can also be translated into a simple network. For this to work out, one may want to assume a neuron that is spontaneously active – that is, a neuron that fires even if not stimulated before. As an alternative, the neuron may receive continuous input from somewhere else.

This neuron, γ, could now receive an inhibitory input from a sensory neuron α so that it would stop firing one time step after α was active. Activity of γ would therefore indicate that α was not active and, thus, that the stimulus to which α responds was not present in the input.

More complex circuits of artificial McCulloch–Pitts neurons are necessary for realizing other logical operations. One may wish to have a neuron that responds to green objects and to round objects, but not to objects that are green and round at the same time. For this purpose, a circuit must be created that realizes the "Exclusive Or" operation. This is not possible based on one single McCulloch–Pitts neuron, but would require an arrangement of such units in two steps, or *layers*.

One possibility, also illustrated in the diagram at the bottom of Figure 6.1, is the following: The input neurons α and β each activate one additional neuron, γ and δ, respectively. In addition, each of them inhibits the neuron the other input neuron activates. Therefore, α activates γ and inhibits δ, and β inhibits γ and activates δ. A fifth neuron, ε, receives activating input from both γ and δ. All thresholds are 1, so that one activating input alone – without presence of additional inhibition – suffices for firing any of the neurons. The ε neuron now is activated if α was active while β was silent (because γ fired) and if α was silent while β was active (because δ fired). Simultaneous excitation of α and β does not effect this circuit because the inhibition caused at the second processing step, the neurons in the middle layer, cancels the effect of any activation. The circuit can be considered a neural analog of an "Either–Or" operation. Activation of the network's cardinal cell, ε, represents the fact that either input feature *a* or input feature *b* was present in the input. It would respond to the orange or to the crocodile, but not to the green ball.

Again, this example illustrates that for certain problems, in this case, the calculation of an "Either–Or" operation, the solution cannot be achieved by one single neuron. It is necessary to introduce processing steps into the network and arrange the computational elements in layers through which activity flows serially. This is an important insight that dates back to the 1940s and 1950s (Kleene, 1956; McCulloch & Pitts, 1943) and was critical for the emergence of the strong interest of neural modelers in multilayer networks in the 1980s and 1990s. A neuron arrangement with an input layer and two subsequent processing steps can realize any event that can be described by a logical formula including the logical operators "And," "Or," and "Not" (Kleene, 1956).

Networks of McCulloch–Pitts neurons can not only be used to detect simultaneous stimuli or events, they can be used to represent complex events spread out in time as well. For example, the sequence of words in a simple

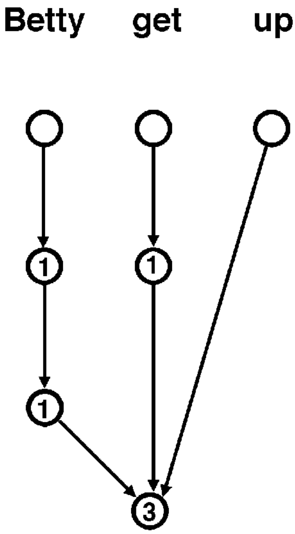

Figure 6.2. Illustration of the representation of a complex event – the word sequence "Betty get up" – in a McCulloch–Pitts network. The neurons in the upper line are called *input units*, each of which is activated by the occurrence of a given word in the input. The neuron at the bottom is a *string detector* that becomes active only if the word sequence occurs. Numbers indicate the threshold of neurons – that is, how many simultaneous inputs they need to become to active. The string detector can be considered a logical unit that indicates that the conjunction of three assertions is correct.

sentence can be represented by a simple network. Assuming neurons that respond specifically to individual words of a language, the McCulloch–Pitts network in Figure 6.2 can be used for detecting the word sequence (1):

(1) Betty get up.

The network shown in Figure 6.2 includes neurons, depicted at the top, that specifically respond to one of the three words. Because they respond to a specific input, they can be called *input units*. At the very bottom of the graph is a neuron that fires if the string of words had been presented and the input units have therefore been fired *in a given order*. Because this neuron at the bottom fires in response to a string of elementary events, one can call it a *string detector*. Apart from the input units and the string detector, the network is composed of neurons whose function is to delay activity caused by the early events so that all relevant information arrives simultaneously at the string detector. Because their function is to delay activation, they can be called *delay units*. This type of network transforms a series of events spread out in time, the serial activation of input units, into simultaneous events, the simultaneous activation of the neurons feeding into the string detector. The string detector in Figure 6.2 has a threshold of 3, thus needing three simultaneous inputs to become active. Therefore, the entire series of three words must be presented to the network to activate the string detector. Missing a word, introducing a delay between words, or changing the order of the words will not allow the string detector to become active.

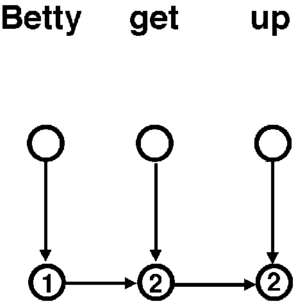

Figure 6.3. Modified McCulloch–Pitts network for sequence detection. Now, sequential activation of the sequence detectors at the bottom would be the criterion for string acceptance.

In these example networks, the criterion for the detection of a complex event by the network was always the activation of one single neuron. This neuron, the string detector in Figure 6.2 and the neurons representing logical operations in Figure 6.1, can be called the *cardinal cells* of the respective networks. Instead of having a cardinal cell in the network, it is also possible to introduce a criterion for event detection based on an activity pattern in which more than one neuron participates. Such a criterion could be that input causes a continuous sequence of neuron firings. The network in Figure 6.3 could then serve the purpose of string detection, because the neurons in the lower row only become active in sequence if one particular input sequence occurs. In this case, several neurons would be involved in the string detection process (bottom row of neurons in Fig. 6.3), apart from the several input units (top row of neurons in Figure 6.3). Delay neurons are no longer necessary, because direct connections between the neurons and activation thresholds include the information about the correct sequence. If string (1), the string "Betty get up," occurs in the input, this causes sequential activation of the three units involved in string detection. If the same words are presented in a different order, at least one of the string-detecting neurons fails to become active.

One may consider most of the networks discussed so far as heterogeneous, because their neurons have different thresholds. Activation thresholds of cortical pyramidal neurons do not differ so much; it is rather the connection strength of synapses that differs widely between neuronal connections. One may therefore argue that these networks are in need of improvement in order to make them more realistic. This could be done by changing all thresholds to 1 and varying connection strengths instead. Because connection strengths between neurons are known to vary, varying connection strengths in artificial networks may be considered more realistic than varying their thresholds. Note that in all diagrams, all connections were either excitatory (lines ending in an arrows) or inhibitory (lines ending in a T shape). If, instead, connection strength was varied, this could be indicated by writing the strength or weight of each link as a number next to the arrow or line. The "Or" circuit in Figure 6.1 would not need to be changed, in this case, the connections all receiving the weight 1. However, the connections in the "And" network would need to be changed to 0.5 so that the cardinal neuron, whose threshold would be set to 1, would need two simultaneous inputs to become active. In the same way, some of the connections in Figure 6.3 would need to be readjusted to 0.5. The function of the circuits would not change as a consequence of these adjustments. The less realistic circuits depicted in the figures can nevertheless be preferred for illustration purposes. One reason for preferring them might be that they require specification of a smaller number of parameters and may, therefore, be considered easier to oversee. Note that in the depicted networks, the number of connections that must be labeled is usually greater than the number of neurons whose threshold must be specified. Depicting the less realistic circuits is also not a problem here because the more realistic equivalent solutions can be derived easily.

Figures 6.2 and 6.3 show two ways in which a sequences of events can be mapped onto neuronal dynamics. In one case, a string detector is present that receives simultaneous input informing it about a specific serial activation pattern. In the second case, the sequence of stimuli is imitated by a sequence of activation of neurons involved in the sequence detection process. As pointed out in Chapters 8 and 9, the two neuronal strategies of string detection that can be postulated on theoretical grounds appear to have analogs in real neuron circuits.

The networks discussed so far can only detect events that occur within a short time interval, usually a few time steps. The temporal extend of these events is certainly always finite. However, it was one of the main points of McCulloch and Pitts's proposal that their networks can also detect events of arbitrary length, endless strings in principle. This can be achieved by introducing loops in the network.

Figure 6.4 presents two neurons that respond to strings of indefinite length. On the left is a neuronal element that becomes active only if it was active at an initial time step, and if it received continuous input since its initial activation. Thus, it is active later only if it was supplied with a continuous sequence of excitatory inputs. As soon as an input fails to stimulate the neuron, its activity immediately ceases, and it is silent until reactivated by a mechanism not specified here. Assuming that the neuron responds to the input "very," it is active if an arbitrarily long word chain including only this particular word has been perceived. One may consider this a somewhat odd

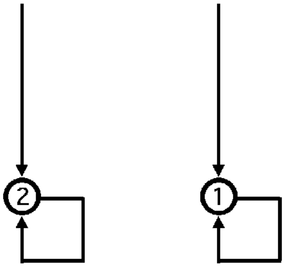

Figure 6.4. McCulloch–Pitts neurons with self-connecting loops. (Left) The neuron with threshold 2 stays active if it was active initially and received continuous input since. Its activity means that a given input was always present (beginning at a predefined starting point). It can represent a string of indefinite length in which a symbol occurs repeatedly. (Right) Activity of the neuron with threshold 1 indicates that a given input has occurred at least once. The latter neuron may be helpful for modeling discontinuous constituents – that is, two morphemes that belong together although they are separated in a sentence.

device, but if the neuron is allowed to respond not to only one particular word, but instead to a class of words, for example, adjectives, its potential usefulness for syntactic processing may become obvious. It could then detect long strings of adjectives as they can occur preceding a noun ("the young, beautiful, attractive, little,... frog"). Because the neuron stays active only during a certain interval if a given input – for example, the word "very" – is continuously present throughout all time steps, one may call it an *"All" neuron*.

In contrast, the neuronal element on the right in Figure 6.4 becomes and stays active whenever the element it responds to has occurred once. Because it continuously stimulates itself once it becomes active, the neuron stays active for, in principle, an unlimited time. Whereas the "All" neuron's activity signals that a particular elementary event has occurred at all time steps (during a certain interval), this present neuron's excitation would mean that at least one time step exists when a particular event was present. It could therefore be called an *"Existence" neuron*. Neurons of these types, expressing the existence and all operations, are necessary to neurally implement aspects of the logical predicate calculus.

If neurons are so helpful in implementing logical calculi, why should they not be helpful in modeling sentences? An "Existence" neuron could be useful for storing the information that a particular word has occurred in the past. The verb *get* and the particle *up* in sentence (1) belong together. When *get* has occurred, it would be advantageous to store this for some time, not only because there may be a delay between the two words as a result of hesitation or breathing but also because other word material may intervene between the verb and its particle. An "Existence" neuron could store the information that the verb has occurred and, via a direct connection, could prime the particle's neural representation so that its activation would be facilitated later.

It is worthwhile to take a closer look at the possibilities opened by the assumption of neuronal units that can, in principle, stay active forever.

Consider sentences (2)–(5):

(2) Betty switched it on.
(3) Betty switched the coffee machine on.
(4) Betty switched the nice brown coffee machine on.
(5) Betty switched the coffee machine she received as a present from her mother at her 25th birthday on.

Here, the distance between the verb "switched" and the particle "on" gets longer and longer. Although examples such as (5) may be unusual in English, one may want to argue that the distance between the two parts of the distributed word "switch . . . on" has, in principle, no upper limit, at least in certain languages (see Chapter 7). Realistically speaking, the distance is seven words or fewer in most cases (e.g., Braitenberg, 1980), but this does not rule out the possibility of longer lags. Distributed words such as "switch . . . on" or "get . . . up" are one type of discontinuous constituent. They illustrate that, in language, sometimes very distant elements are in a close relationship, a phenomenon called *long-distance dependency* (see Chapter 7). To produce or understand sentences such as (5), it is therefore necessary to be prepared to store the earlier of the mutually dependent elements or the necessity that the second is required, and there is perhaps no upper limit for the time this fact must be kept in mind. This issue will be discussed in depth in Chapters 7 and 10.

Clearly, the task to store the piece of information that the verb stem has occurred or that a verb particle is required can be carried out by a bistable unit that is activated and stays active thereafter. One possible realization in a neuronal network model is presented in Figure 6.5. Here, the "verb neuron" that projects onto the "particle neuron" keeps itself active because of the self-activating loop. Thus, the network (or fraction of a network) detects a string consisting of an initial verb and a particle occurring after

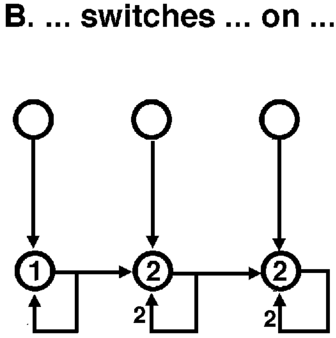

Figure 6.5. (Top) A network that shows sequential activity of the sequencing unit (lower line of neurons) if the string "Betty switches . . . on" appears in the input. The delay between the two parts of the discontinuous constituents can vary. The "2" next to the arrow forming a loop indicates connection strength. (Bottom) By extension, each sequencing unit can be assumed to be *bistable*, so that delays would be possible between all input symbols. This may be useful, for example, to allow for introduction of a relative clause after the noun.

the verb, with, in principle, no upper limit for the time delay between the two.[1]

This example of the representation of two linguistic elements that belong together but are nevertheless separated in time or space shows that networks of McCulloch–Pitts neurons can solve linguistic problems that cannot be solved by probabilistic grammars such as Markovian sources (Charniak, 1993; Markov, 1913). Markov chains and other so-called k-limited stochastic models use the probabilities with which a $(k+1)$th element follows a string of k other elements (obtained from large text corpora) to calculate the probability with which a particular word (or letter) follows a given sentence fragment.

An example how this works is the following: Markov (1913) used Pushkin's novel *Eugene Onegin* to calculate the probabilities of vowels and consonants to follow each other. He also calculated the probabilities of consonants (C) and vowels (V) to follow series of CC, CV, VC, and VV. In this case, $k = 2$, that is, only two preceding elements were used to calculate the conditional probability of the next letter type. The same can be done for words in sentences and for larger k' s (for further discussion, see Miller & Chomsky, 1963; Shannon & Weaver, 1949).

However, the fact that an element occurring in a string (e.g., the root "turn") leads to a 100 percent probability of occurrence of another element (the verb particle "on") at an indefinite time step later cannot be modeled in the Markovian framework. To model it in the Markovian framework, the two distant elements that belong together would have to be treated differently from the intervening string. Both sentence (6), which includes a verb particle, and sentence (7), which includes an intransitive verb without a particle.

(6) Betty *steht* morgens um halb drei Uhr morgens mit einem mürrischen Gesicht *auf*. (including the particle)
(7) Betty *geht* morgens um halb drei Uhr morgens mit einem mürrischen Gesicht. (without particle)

Nevertheless, the ten words preceding the particle in sentence (6) are the same as the ten words at the end of sentence (7). Therefore, the difference between the two would escape a k-limited stochastic source if $k = 10$. The probability for the particle to occur is primarily influenced by the distant morpheme (with no strict upper limit for the distance) – not by its

---

[1] This preliminary solution raises additional problems. First, the criterion for string acceptance must be modified. Second, the bistable element must be switched off after the particle occurs in the input. In later chapters, in particular 9 and 10, related ideas will be looked at in more detail.

next neighbors. To model this, probabilities would have to be calculated for strings of arbitrary length. An extension of the Marcovian approach is necessary – for example, so-called Hidden Marcov models – to capture this property of sentences. A McCulloch–Pitts network seems to be able to solve one serial-order problem in language that stochastic models of a particular type have systematic difficulties with. One may ask in which way a McCulloch–Pitts-based approached could help in capturing serial-order relationships in language, where its principle limitations are, and how these can be overcome.

McCulloch–Pitts's framework allows for modeling logical operations at the level of neurons. It can also be used to model serial order of complex spatiotemporal events. Historically, McCulloch and Pitts's proposals were crucial for the emergence of automata theory, without which the development of modern computers would have been unthinkable. In language science, they inspired research on grammar algorithms. Grammar models such as finite state grammars and augmented transition networks were built on the proposed neural mechanisms. These approaches to serial-order mechanisms were successful in modeling particularities of grammatical rules (see, for example, Winograd, 1983), although the need to incorporate more sophisticated rules sometimes led to extensions that made the resulting devices somewhat more difficult to relate to neuronal structure. Examples are the recursive embedding of networks and the addition of registers that are acted on under special conditions whose implementation in terms of neurons may appear opaque. This is not to say that these entities and processes could not themselves be transformed into circuits of artificial neurons. However, in the present context, the task is to provide candidate circuits for syntactic processing that may be realized in the brain. The circuits originally proposed by McCulloch and Pitts seem to be closer to this general aim than some of the more recent and more sophisticated syntactic algorithms rooted in their work.

6.2 Symbolic Connectionist Models of Language

The idea to model language and cognitive processes on the basis of artificial neuron-like devices was applied at different levels. Syntactic models, such as finite state grammars, were proposed that modeled aspects of serial-order mechanisms, and at the level of linguistic units – words and language sounds, for example – processing models were proposed that specified the activation processes envisaged to occur during production and comprehension.

According to modular approaches to cognitive psychology, the mental language processor consists of quasi-autonomous subprocessors or modules, and therefore language processing is being considered the result of subprocesses carried out by these. The subprocesses envisaged to be involved in language comprehension are, for example, input-feature analysis, letter or phoneme analysis, word-form processing, and semantic analysis (Ellis & Young, 1988). A similar but reverse cascade has been assumed for the putative subprocesses of language production that finally results in movements of the articulators or the (writing) hand (Garrett, 1980). The postulated subprocesses of word comprehension and production are assumed to occur sequentially or in an overlapping cascaded manner. It cannot be disputed that during certain types of language production and comprehension, information about stimulus features, letters or phonemes, word forms and semantics are being processed. One may therefore refer to different levels or stages of language processing, such as the word form or lexical stage or the semantic level. A specific proposal immanent to one class of modular models now is that information processing is autonomous at each of the different levels.

In symbolic connectionist models, the subprocessors of modular models have been replaced by layers of neuron-like elements, the assumption being that individual artificial neurons or *nodes* represent acoustic or visual features, phonemes or graphemes, word forms, and word meanings (Dell, 1986; Dell et al., 1997; McClelland & Rumelhart, 1981). Figure 6.6 presents one model (Dell, 1986) in which each representation of a word is composed of a central lexical node and its associated meaning-related (semantic) and form-related (phonological) nodes. In this network, a stimulus word first activates the phonological nodes corresponding to its sounds. These would, in turn, strongly activate the lexical node of the respective word and partially activate lexical nodes representing words phonologically similar to the stimulus word. Activity then spreads to the semantic layer. This model has reciprocal connections so that backward flow from the semantic to the lexical layer and from the lexical to the phonological layer would also be allowed. Computations at the different levels of the network can take place at the same time and there can be a continuous exchange of information between levels during computation. Each neuron of the network would at each time step compute the sum of its inputs and receive an activity value accordingly. In contrast to McCulloch–Pitts neurons, the networks' neurons would not need to have a fixed activation threshold, but their activity level may vary continuously. A special criterion, for example, which node is most active in the lexical layer, may serve as the criterion for considering the word it represents to be selected among alternative words whose nodes have also been activated to some degree. The nodes are assumed to stay active for some time after their stimulation, not only because they continuously send

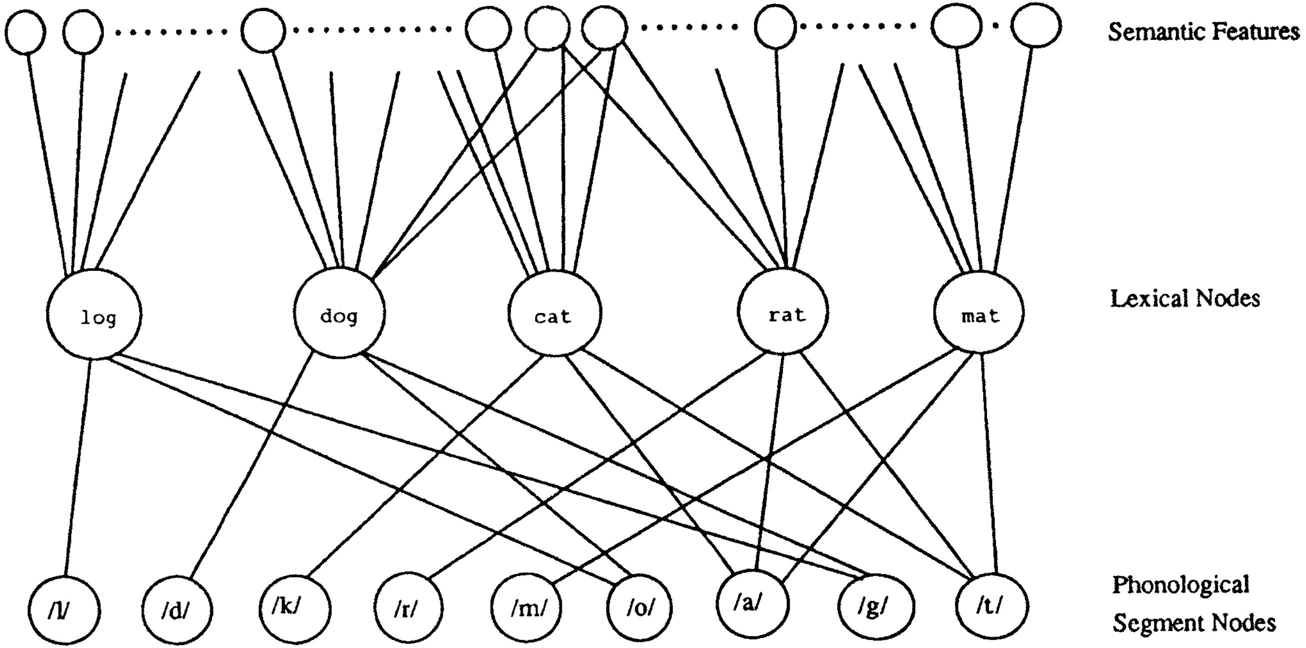

Figure 6.6. A model of word processing proposed by Dell (1986). Each word is represented by a lexical node that is connected to a specific set of phonological and semantic nodes representing speech–sound related and meaning related features of the word, respectively.

out and receive activity from their associated nodes in other layers, but also because each node is assumed to hold back some of its activity by an internal memory mechanism.

The model has been applied successfully to various psycholingusitic and neurolinguistic data sets and can, for example, imitate patterns of speech errors made by normal speakers and aphasic patients. The network can imitate *semantic errors*, replacements of a word by another one close in meaning. In the model, semantic errors are based on activity flow between lexical nodes through their common associated semantic nodes. *Phonological errors*, replacement of words with similar sounds, can be based on phonological nodes strongly associated with more than one lexical node. The model also explains why errors that are both phonologically and semantically related to the target word are far more frequent than one would predict on stochastic grounds (Dell, 1986; Dell et al., 1997).

Different authors diverge in the exact architecture postulated for their symbolic connectionist networks, with some proposing that more node types, layers, and levels of processing are necessary. One influential proposal is that apart from phonological, lexical, and semantic representations, there should be a separate representation of a word's syntactic properties, and that this syntactic word node should be introduced between the lexical and the semantic layers (Levelt, Roelofs, & Meyer, 1999).

A major point of divergence between proponents of the modular and the interactive approaches to cognitive processing concerns the processing properties of the systems of nodes and connections. As mentioned, the modular approach states that processing at each level is autonomous – that is, not influenced by processes occurring at the same time at other levels – whereas the alternative view is that there is continuous forward and backward activity flow. Furthermore, the strict modular view puts that activity flow is possible only in one direction, therefore necessitating separate networks for modeling speech production and perception. The discussion between modularist and interactionist schools was quite vivid, with the outcome being that both approaches can provide models for impressive data sets (Dell et al., 1997; Norris, McQueen, & Cutler, 2000).

From a neuroscientific perspective, some assumptions immanent to modular models may appear difficult to anchor in brain matter. For example, the idea that activity should be allowed to flow in only one direction through a network is in contrast with neuroanatomical observations that most of the corticocortical connections are reciprocal (Pandya & Yeterian, 1985; Young et al., 1995). Also, the idea that the processing of different types of linguistic information – phonological, lexical, and semantic, for example – should take place without interaction would not have a necessary correlate in, for example, an associative memory model of word processing. In such a model, strong links would be present, for example, between neurons responding to acoustic features of a stimulus word and those sensitive to visual features of objects to which the word refers. These links would allow for fast information exchange between phonological and semantic parts of the distributed word representation. Having said this, it should be added that this by no means implies that a modular approach would be incompatible with a neurobiological perspective.

The interactive perspective discussed here comes close to a cell assembly model as proposed in Chapter 4. In both types of models, semantic and phonological information about a word is linked through strong connections. In both model types, an activated representation maintains its enhanced activity status for some time. The representation of a word is the union of neural elements spread out over different "layers," or brain areas. There is one important difference between the models, however. The connectionist architecture includes nodes that represent and process only one lexical entry. These word-specific neural elements are not necessary in the cell assembly proposal. The coactivated neuron sets in primary sensory and motor cortices are assumed to be held together by neurons in other areas, but neurons involved only in the processing of one particular word would not be required. Each neuron included in the functional web representing one particular word could also be part of cell assemblies of other words as well. It is clear that, in a computational architecture, word-specific nodes make the simulations easier, because the activity level of the lexical node and its relation to that of other lexical nodes provides a simple criterion for selecting one of the words. In an associative network of overlapping functional webs, it may be much more difficult to determine which of the assemblies is most active. Furthermore, activity excess is one of the dangers an associative memory model runs into, a problem that is easily avoided in the symbolic network: for example, a regulation mechanism within a lexical layer only allowing one node to be selected could be used. Therefore, the cell assembly model requires a regulation mechanism as well (Braitenberg, 1978a; Wickens, 1993). The regulation issue was addressed in Chapter 5.

Although the lexical layer may have computational advantages, one may argue that such a layer of word-specific neural element is not an *a priori* requirement within a neurobiological system. The functional web approach would suggest that there are neuron sets in different primary cortical areas that are activated, on a regular basis, when a given word is being processed, and that there are neurons in other areas providing the connections between these neurons. These additional neurons whose main function is to hold the functional web representing a word together would respond to excitation of sensory and motor cells also being part of the ensemble. If each of these *binding neurons* was also be part of other cell ensembles, this may even have advantages from a neurocomputational point of view. As mentioned in Chapter 5, simulation studies of associative memories showed that the best storage capacity is achieved if neuronal assemblies are allowed to overlap, so that each neuron is part of several assemblies (Palm, 1982). Therefore, it is not necessary to assume word-specific neurons in a network in which each word has its specific neuron set. However, these considerations do not exclude the possibility that in a network of overlapping neuronal assemblies, each representing one individual word, there might be neurons that are part of only one assembly. Again, however, these cardinal cells (Barlow, 1972) would not be required. Still, an important common feature between an approach postulating functional webs and a symbolic connectionist network is that each word is assumed to be represented as an anatomically distinct and functionally discrete set of neurons or neuron-like elements.

6.3 Distributed Connectionist Models of Language

The distributed network type most commonly used to model language resembles symbolic networks because both network types are made up of layers of neurons and connections between layers. An important difference is as follows: Most symbolic networks include so-called local representations – usually single artificial neurons – that can represent elementary features of the in- and output, but also more complex entities; for example, letters, phonemes, and words. An example already discussed is the separate lexical node postulated for each individual word. In contrast, distributed networks use activity vectors specifying the activity of all neurons in one or more layer(s) to represent words and other more complex linguistic structures. In this view, all neurons contribute to the processing of every single word. One cannot say that only the active neurons of an activity vector would be relevant for coding the respective word because changing the value of one active neuron in the vector (to inactivity) and changing that of an inactive neuron (to activity) may have a strong effect as well. This is particularly so if the coding is not sparse, as is frequently the case in the so-called hidden layer of multilayer networks. Thus, each neuron contributes to the fully distributed representation. One may therefore conclude that distributed connectionist networks and localist symbolic networks reflect two extreme views at the two ends of a continuum. Each individual neuron would be a “Jack of all trades,” so to speak, in fully distributed approaches, but a master of almost nothing (only one single word or phoneme, etc.) in the localist framework.

It is clear that both views could be correct descriptions of what actually is the case in the brain, but it is also clear that there is room for alternative proposals in between the extremes.

The cell assembly framework offers such an alternative possibility between these extremes. Each neuron is envisaged to contribute significantly to the processing of some words, but not necessarily to all of them. Each neuron has a critical role in distinguishing between some lexical alternatives, but may nevertheless be irrelevent for distinguishing between words to which it corresponds. Not one cardinal cell and not the entire system would process a single word, but a strongly connected neuron set, a functional web. A lexical representation would be the union of neurons involved in processing the motor program, acoustic properties, and semantic features of a word.

A further difference between a cell assembly model on the one hand and a fully distributed approach or a localist account on the other hand lies in the structure of the assumed between-neuron connections. The distributed connectionist networks most commonly used in language simulations, and also their localist sisters, do not usually include direct excitatory connections between the neurons in one layer. For example, an input layer in which each neuron represents a phonological feature would leave these feature neurons without direct connections, although their activity may, in the absence of other active neurons, define a word. In contrast, a model of functional webs based on neuroanatomical data must assume strong links within cortical areas and between directly adjacent areas, and therefore between neurons related to sensory processes in a certain modality. This implies that the different neurons sensitive to acoustic properties of a stimulus word have a relatively high probability to be connected to each other directly, rather than by way of intervening areas or layers. Because the active neurons defined by an activity vector are not connected to each other directly, they do not form a functionally coherent system. This distinguishes them from cell assemblies or functional webs with strong internal links between their neuron members, especially between their closely adjacent neurons. However, given there are reciprocal connections between the layers of the network, artificial neurons in different layers may strengthen their connections to neurons in other layers as a consequence of associative learning, thereby indirectly linking the active neurons defined by individual activity vectors. Still, the structural differences between the connectionist architectures and a view incorporating more details about neuroanatomical connnections remain.

The classical type of distributed connectionist network is called the *perceptron* (Minsky & Papert, 1969; Rosenblatt, 1959). It consists of neurons arranged in two layers, the input and the output layers. Each neuron in the input layer is connected to each neuron in the output layer. Activity flows from input to output only. Connection strengths (or weights) vary as a function of associative learning between pairs of input–output patterns. Rosenblatt (1959) proved that perceptrons can learn to solve only a certain type of simple classification problem called *linearly separable*. Rumelhart, Hinton, and Williams (1986) showed that a modification of the perceptron's architecture, through the addition of one additional layer of neurons, the so-called hidden layer between the input and output layers, and by an extension of the learning rule, this limitation can be overcome. The networks were now able to learn to solve more complex classification problems, for example, classifications requiring the Either–Or operation, and this made them much more interesting for cognitive scientists. It was mentioned in Section 6.1 that a network of a certain type requires three neuronal steps to solve the Either-Or problem (Kleene, 1956).

Three-layer perceptrons have been used successfully to model aspects of language processing. For example, there are models that classify speech signals (Waibel et al., 1995), others that mimic important aspects of the infant's learning of language specific information as described by elementary rules (Hare, Elman, & Daugherty, 1995), and simulations of the effects of focal brain lesions on language functions (Hinton & Shallice, 1991; Plaut & Shallice, 1993) and of the recovery of language functions after stroke. The results of these simulations are not only of great theoretical relevance, but they have found many useful practical applications as well.

Two- or three-layer distributed connectionist architectures do not have memory. An input pattern activates the input layer and activity spreads to the (hidden and) output layer(s). This network architecture must be modified for addressing problems posed by syntax. To allow the network to process long strings of words, it was necessary to introduce a memory mechanism in the architecture that allows the storage of information about past events for a longer time span. Such a memory is necessary for assessing syntactic relationships between temporally distant language units, as, for example, between the first and last words of a long sentence. Therefore, a *memory layer* allowing for reverberation of activity, thereby providing the basis of information storage, was added to the three-layer architecture (Elman, 1990). Either the hidden layer was given within-layer connections so that each of its neurons fed back onto itself and the other neurons in the hidden layer, or loops were introduced by connecting the hidden layer reciprocally to a separate memory layer. Compared to three-layer perceptrons, these networks including a memory layer could be shown to be more powerful in storing serial-order relationships. They are capable of learning subsets of syntactically complex structures, for example, aspects of so-called center-embedded sentences (Elman, 1990; Elman et al., 1996).

6.4 Hot Topics in Neural Network Research on Language

Among the many facets of connectionist research on language, work on two topics can nicely illustrate in which way our understanding of the language mechanisms have been furthered by simulation studies. The point here is not so much that the actual simulation taught us important lessons. The critical colleagues who consider many of the results of simulation studies as trivial and emphasize the relevance of rigorous thinking about a problem, after which a simulation might be close to redundant, may therefore be partially correct. The point here is rather that simulation studies and comparable theoretical work made scientists think about details of neuron circuits, and this led to important insights.

Both issues to be addressed in this section are related to deficits in language processing, seen either in aphasia or in young infants before they are fully capable of speaking their mother tongue. One issue is the explanation of deficits in processing words of certain types. The other relates to young infants' acquisition of what one may want to describe as *syntactic rules*.

6.4.1 Word Category Deficits

Certain types of aphasia are surprising because the patients can produce words of one type, but are unable to produce words of a second type. This issue is addressed in Chapter 4 along with neuroimaging results. The principle observation that a word kind can be affected selectively by brain lesion has first been made with regard to a disturbance called *agrammatism* (Pick, 1913). Agrammatic patients have difficulty producing words primarily characterized by their grammatical function, such as articles, pronouns, auxiliary verbs, prepositions, and also inflectional affixes. Geschwind (1974), a neurologist, recommended basing the diagnosis of agrammatism on a simple test: Ask the patient to repeat Sentence (8).

(8) No ifs, ands, or buts.

An agrammatic patient would usually not be able to repeat the sentence, and would otherwise not be able to use the words included in it correctly. The affected words are used mainly as grammatical function words and do not refer to objects or actions. Difficulties with the respective word forms are present even if these words are not actually used in their normal grammatical function but instead, for example, in their nominalized forms (as nouns), as

Sentence (8) illustrates. The abstract function words whose meanings are not imageable at all are sometimes not the only word kinds affected. For some patients, abstract nouns and verbs whose meaning is difficult to imagine are also difficult to produce. A deficit in producing function words and other low-imageability words can be most pronounced in language production but can be present in language perception as well (Pulvermüller, 1995; Pulvermüller, Sedat et al., 1996). A deficit most pronounced for abstract low-imageability words is also common in a neurological reading deficit called *deep dyslexia* (Marcel & Patterson, 1978) and in *semantic dementia* (Patterson & Hodges, 2001). One view on agrammatism and similar disturbances is that the imageability of words is a crucial factor in determining their availability after brain lesion. Low-imageability words are more difficult than comparable high-imageability words.

Even more surprisingly, however, agrammatism has a "mirror image" syndrome called *anomia*. Anomic aphasics do not have difficulty using function words and may even produce complex sentences without any problem but cannot find some well-imageable content words from the lexical categories of nouns, adjectives, and verbs (Benson, 1979).

How can the double dissociation between the deficit for low-imageability words and function words in agrammatism and the deficit for the more concrete and imageable content words in anomia be explained? In the cell assembly framework discussed in Chapter 4, a highly imageable word would be represented by a perisylvian cell assembly with strong connections to neurons in other brain areas organizing referential aspects of the word, that is, the motor programs and perceptual patterns to which it can refer. The cell assembly would therefore be distributed over wider cortical areas, including perisylvian areas but also other sensory- and action-related areas as well. In contrast, a function word lacking any concrete associations that cannot be used to refer to concrete objects or actions would be represented by a functional web without strong links to neurons in action- or perception-related areas. Therefore, function words and also highly abstract and not imageable content words would be realized in the brain by more focal functional webs restricted to the perisylvian cortex. Figure 6.7 illustrates schematically the postulated difference in circuits in the language-dominant hemisphere. A further implication would be that the function words' networks are more strongly lateralized than are the networks representing highly imageable content words. The word types are thus assumed to be realized as functional webs with different cortical topographies (Pulvermüller, 1995).

Assuming more localized perisylvian functional webs for grammatical function words and more widely distributed cell assemblies for concrete nouns, verbs, and adjectives makes a few additional comments necessary.

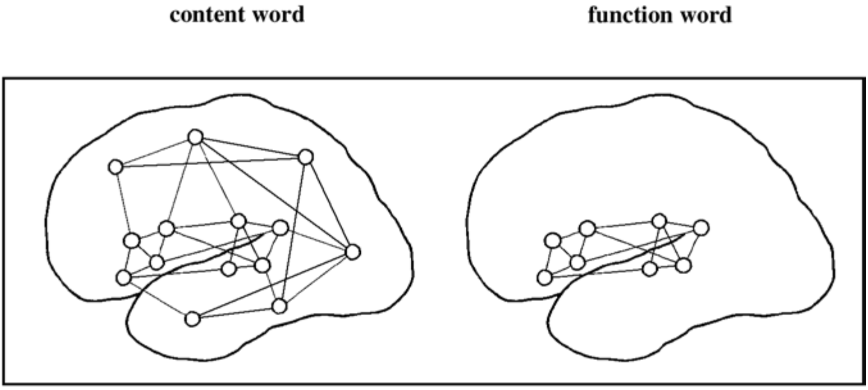

Figure 6.7. Schematic illustration of left-hemispheric distributions postulated for cell assemblies representing high-imageability content words (left) and highly abstract function words (right).

First, it may appear more adequate to postulate such differential representation not at the level of individual words, but at that of meaningful units, morphemes, instead. In this view, not entire content words, but "content morphemes" or the stems of content words, are represented (cf. Section 4.4). (The reader may forgive the imprecision that the more common term, *content word*, is usually used in this book because it is much more common.) Furthermore, there is no principle difference in the grammatical function of a function word, such as the article "the", and a functional affix, such as the plural suffix "-s" attached to nouns. Their inclusion in a sentence usually does not add semantic information to it, but may make the word string either acceptable or unacceptable. The neuronal representations of function words and affixes could therefore exist without a semantic part. At least, they can be conceptualized without a part including information about referential meaning that would be stored primarily outside the perisylvian region. In this respect, the functional webs representing different kinds of *function items*, words and affixes, may be similar.

A second point one may wish to raise is the following: The assumption that no semantic information is being added to a sentence by the inclusion of function items does not generally hold true for all members of this category. The regular past suffix "-ed" and the auxiliary verb form "was", for example, include information about time, and it is clear that a cortical representation of these lexical items must include this information as well and bind it to the phonological information about the respective form. However, it is less clear where in the brain this semantic information characterizing some function words and affixes should be laid down. This information is not about these words' referential semantics, and, therefore, considerations such as the ones made in Chapter 4 are not relevant. By postulating cell assemblies, or functional webs, for function words and affixes, it remains clear that the semantic information immanent in some of them must have a brain internal correlate, although it is doubted that this locus of storage and processing is cortical and outside the perisylvian areas. In this context, it is relevant that the differential cortical processing of function words and stems of concrete content words receives support from neurophysiological imaging studies (Brown, Hagoort, & ter Keurs, 1999; Neville, Mills, & Lawson, 1992; Nobre & McCarthy, 1994; Pulvermüller, Lutzenberger, & Birbaumer, 1995; Shtyrov & Pulvermüller, 2002a). In the context of studies reviewed in Chapter 4, these neurophysiological differences are best explained in terms of the difference in referential semantics between content words and function items (for discussion, see Pulvermüller, 1999).

A parsimonious explanation of the double dissociation between agrammatism and anomia becomes possible on the basis of the proposal that function items are processed by strongly lateralized neuron ensembles restricted to the perisylvian cortex, whereas content words have less lateralized corresponding neuron ensembles distributed over various areas of both hemispheres. Presumably, the likelihood of a word to be affected by a brain lesion depends on the degree to which its word representation has been damaged.

A lesion restricted to the perisylvian regions, for example in Broca's region in the inferior frontal lobe, would remove a high percentage of neurons included in a function words' webs. The concrete content words' webs that have their additional neurons outside perisylvian space (and also in the other hemisphere) would not suffer so much, that is, a smaller percentage of their neurons would be affected by the same lesion. This explains deficits primarily affecting function words and other abstract items, as seen in agrammatic patients and related syndromes. The mirror-image pattern of deficits would be expected if a lesion affected areas primarily outside the perisylvian region, for example, in the inferior temporal lobe, the temporooccipital region, or various areas in the hemisphere not dominant for language. In this case, only the concrete content words' webs would be affected, thereby explaining a selective deficit in processing these items. This explanation of the double dissociation between agrammatic and anomic deficits has been grounded in neural network simulations and is in good agreement with data about the cortical locus of lesions causing these disturbances (Pulvermüller, 1995; Pulvermüller & Preissl, 1991). Focusing on deficits for low- vs. high-imageability content words in deep dyslexia, a similar account has been developed based on distributed connectionist networks (Hinton & Shallice, 1991; Plaut & Shallice, 1993).

Neural network accounts of category-specific deficits for words and concepts may be grounded in a standard connectionist architecture, a three-layer perceptron, for example. A selective deficit for one word and conceptual kind may then relate to properties such as the frequency with which the words are being used, or the amount of semantic feature overlap the words – and their neural representation – share with other words from the same category. For modeling more fine-grained category-specific deficits, as mentioned in Chapter 4, it might be advantageous to use networks that include subcomponents specifically handling information processed in different brain areas. One way to go is to introduce components for action-related, visual, and other sensory information types. The word and conceptual representations could be laid down as distributed but strongly connected neuron sets differentially distributed over some of the network's components. Category deficits would then result from lesions in one of the network parts (Farah & McClelland, 1991; Humphreys & Forde, 2001; Plaut & Shallice, 1993; Pulvermüller & Preissl, 1991).

6.4.2 The Development of Rules in the Brain

The modeling of rule-like verbal behavior is another illustrative example for the fruitfulness of neural network research on language. It is sometimes assumed that symbolic algorithms are necessary for explaining the behavior described by linguistic rules and that a description in terms of rules is incompatible with a neural network approach. For producing a past tense form of English, one would, accordingly, use an abstract rule such as the following addition scheme:

(9) Present stem + Past suffix = Past tense form

An algorithm of this kind could model the concatenation of the verb stem "link" and the past tense suffix "ed" to yield the past tense form "linked," and, more generally, it could be used to derive any other regular past form of English. However, it is difficult to see how an irregular verb such as "think" or "shrink" could yield a past tense form by a similar rule. In the extreme, one would need to assume rules for individual words to provide algorithms that generate, for example, "went" from "go." This would require stretching the rule concept, and linguists have therefore proposed that there are two distinct cognitive systems contributing to language processing: a symbolic system storing and applying rules and a second system storing relationships between irregular stems and past tense forms in an associative manner (Pinker, 1997).

From the perspective of neural networks, however, one may ask whether two separate systems for rules and exceptions are actually necessary to handle regular and irregular inflection. Rumelhart and McClelland (1986, 1987) showed that a two-layer perceptron can store and retrieve important aspects of both past tense rules and exceptions. It can even produce errors typical for children who learn past tense formation, such as so-called overgeneralizations (e.g., "goed" instead of "went").

From a linguistic perspective, the two-layer model of past tense proposed by Rumelhart and McClelland has been criticized because it did not appropriately model the fact that rule-conforming behavior is by far most likely to be generalized to novel forms. The past tense form of a newly introduced verb such as "dif" will thus almost certainly receive an "ed" ending if one intends to use it in the past tense ("diffed"). This is even so in languages in which most verbs have irregular past tense forms and only a minority of the verbs conform to the rule. The rule is nevertheless used as the default, and generalized to novel forms and even to rare irregular items. This is a problem for a subset of connectionist models because the strongest driving forces in associative networks are the most common patterns in the input.

However, there are distributed three-layer networks that solved the problem of default generalization surprisingly well (Hare et al., 1995). An important determinant was that rule-conforming input patterns are maximally *dissimilar* whereas the members of an irregular class resemble each other. Consider the different regular forms *watch*, *talk*, and *jump* in contrast to the similar members of an irregular class *sing*, *ring*, and *sting*. Because the regulars are so heterogeneous, they occupy a wide area in input space. The representation in input space of a novel word is thus most likely to be closest to those of one of the many different regular forms, and this is one reason why so many new items are treated as regular by the network. However, if a newly introduced item happens to strongly resemble many members of a regular class, for example, the pseudo-word *pling*, it is, in many cases, treated as regular. These observations may lead one to redefine one's concept of regularity: A rule is not necessarily the pattern most frequently applied to existing forms, but is always the pattern applied to a most heterogeneous set of linguistic entities. The heterogeneity of the regular classes may explain default generalization along with the great productivity of rules.

The simulation studies of the acquisition of past tense and other inflection types by young infants suggest that neural networks consisting of one single system of layers of artificial neurons provide a reasonable model of the underlying cognitive and brain processes. In this realm, the single system perspective appears equally powerful as an approach favoring two systems: one specializing in rule storage and the other in elementary associative patterns.

Neuroscientific data and theories have shed new light on the issue of a single-system vs. double-system account of rule-like behavior. The discovery of patients with brain lesions who were differentially impaired in processing regular and irregular past tense forms was important here. Patients suffering from Parkinson's disease or Broca's aphasia were found to have more difficulty processing regular forms, whereas patients with global deterioration of cortical functions as seen, for example, in Alzheimer's disease or semantic dementia showed impaired processing of irregular forms (Marslen-Wilson & Tyler, 1997; Ullman et al., 1997). This double dissociation is difficult to model using a single system of connected layers, but is easier to understand if different neural systems are used to model regular and irregular inflection.

Another argument in favor of a double system account comes from a neurobiological approach. Within the cell assembly framework, a typical verb stem has its putative neurobiological correlate in a widely distributed functional web that includes neurons involved in the processing of actions and perceptions related to the meaning of the verb. The to-be affixed or infixed phonological material – for example, the ending "ed" or the [æ] inserted into "ring" may initially be processed by a network restricted to perisylvian space. The widely distributed assembly and the more focal perisylvian assembly can be linked by two types of connections. One possibility is provided by local connections within perisylvian areas, and the other by long-distance connections between the verb stem's assembly neurons far from the Sylvian fissure and the perisylvian neurons contributing to the processing of the past tense form. The local and long-distance links have different likelihood, two adjacent neurons having a higher probability of being connected than two neurons in distant cortical areas (see Chapter 2). Assuming that most of the relevant connections between the two cell assemblies are indirect, that is, through one additional neuronal step, the situation can be schematized as in Figure 6.8. Two systems of connections, a high-probability pathway linking perisylvian neurons and a low-probability pathway between neurons outside perisylvian space and neurons inside, would be available for storing links between present and past tense forms. Given parameters are chosen appropriately, the two low- and high-probability systems specialize differentially in the storage of rules and irregular patterns. Similar to a two-layer perceptron, the low-probability system is best at storing the simple mapping

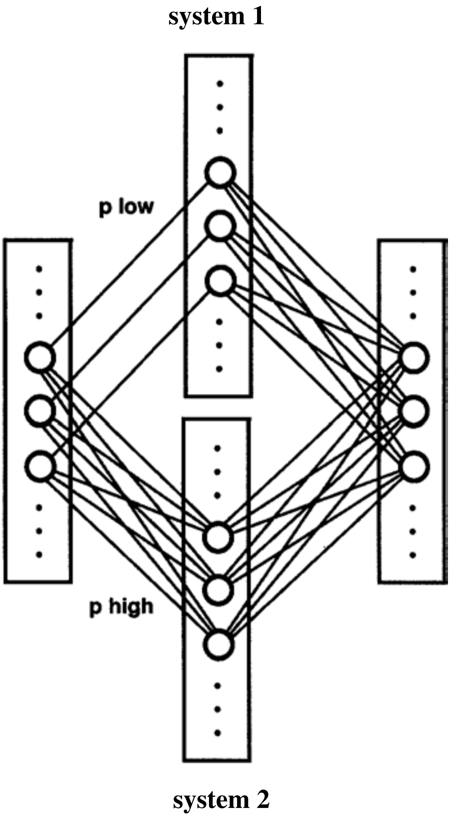

Figure 6.8. A neuroanatomically inspired network architecture consisting of two systems of connections specializing in exception learning and rule storage, respectively. The system at the top includes a link with low connection probability. The system at the bottom has high connection probabilities throughout. Associative learning in this system leads to rule storage primarily in the high-probability system (bottom) and to exception storage primarily in the low probability system (top). There is reason to assume that two types of neuroanatomical links are relevant for the learning of morphology. (For further explanation, see text.) Adapted from Pulvermüller, F. (1998). On the matter of rules. *Network: Computation in Neural Systems*, *9*, R1–51.

between irregular present forms that resemble each other and their past tense forms. In contrast, the complex classification problem, the mapping between the heterogeneous regular stems and their past tense forms, is best accomplished by the three-layer component with high connection probabilities. A full three-layer architecture is necessary for this, because Either-Or relations are required for the classification of the heterogeneous items in the regular class. When the two components are damaged differentially, the network produces the double dissociation between regular and irregular inflection seen in neuropsychological patients. This approach explains the neuropsychological double dissociation while being consistent with aspects of the acquisition of past tense formation by young infants (Pulvermüller, 1998). The explanation is based on principles of cortical connectivity and has implications for the cortical loci of the wiring organizing rules and exceptions, respectively.

Together, the neuropsychological double dissociation and the neurobiological consideration argue in favor of a two-system model of regular and irregular inflection. In contrast to a modular proposal stating that two systems are concerned exclusively with regular and irregular processes, respectively, the neuroscientific variant would suggest a gradual specialization caused by differential connection probabilities. The fruitful debate between scientists favoring single- or double-system accounts of rule-like knowledge reveals the importance of multidisciplinary interaction between the linguistic, cognitive, computational, and neurosciences.

CHAPTER SEVEN

Basic Syntax

This chapter highlights a few problems of serial order that may pose problems to a biologically realistic model, problems best explained in the context of theories designed to solve them. Therefore, the following paragraphs sketch examples of linguistic descriptions of syntactic regularities. In addition, technical terms are introduced. Occasionally, a sketch of what a neuron-based approach to syntax might look like may intrude but is developed systematically only in Chapters 9–12.

There are many different approaches to syntax in theoretical linguistics, and it is not necessary to give an overview of them in the present context. This chapter highlights examples, their choice being primarily motivated by the historical development. So-called *phrase structure grammars* and their offspring rooted in the work of Harris (1951, 1952) and Chomsky (1957) are in the focus, because the superiority of phrase structure grammars to a model of serial order in the McCulloch–Pitts (McCulloch & Pitts, 1943) tradition was one of the main reasons in the 1940s and later to base syntax theories on these more abstract algorithms rather than on neuronal algorithms. Apart from approaches related to and building on phrase structure grammars, an alternative framework whose roots also date back to the 1940s is mentioned occasionally. This framework, or family of theories, goes back to Tesnière and is called *dependency grammar*. Clearly, phrase structure grammars and dependency grammars have been modified and much developed, but, as is argued later in this chapter, some of their underlying main ideas persist.

In addition to an introduction of syntactic terminology and theory, this chapter reviews arguments in favor of and against different views on serial-order mechanisms. Here, the main focus is on the question of what abstract syntactic theories achieve more than neuron-based proposals. It is also asked what the neuron-based proposals could offer what the abstract syntactic theories lack.

A linguist familiar with most syntactic theories may safely skip most of this chapter, but may want to glance over the list of questions that must be addressed when building on a McCulloch–Pitts approach to syntax.

7.1 Rewriting Rules

Syntactic regularities are sometimes expressed in terms of *rewriting rules*. A rewriting rule is a formula expressing that something can be replaced by, or be rewritten as, something else. The rule consists of an arrow and symbols to its left and right. The arrow can be read as "is rewritten as" or "is replaced by." A subtype of such rules, *context-free rewriting rules*, have only one single symbol to the left of the arrow and one or more symbols on its right. Examples of context-free rewriting rules are given later. A set of context-free rewriting rules defines a *context-free grammar*. A set of rewriting rules can also be called a *phrase structure grammar*, or PSG.

Calling a set of rewriting rules a *grammar* implies a formal view on what a grammar should actually achieve. This view restricts grammar to aspects of the form of sentences, without consideration of the meaning of words and strings and their role in communicative interaction. In more recent proposals, the interaction of formal syntactic rules and principles with semantic processes is also considered. Although it appears justified to restrict consideration to the question of how serial order of meaningful language units is achieved, the task of a grammar may nevertheless be considered to cover not only the form of sentences, but aspects of their meaning as well.

In modern linguistics, great breakthroughs have been achieved in the formal description of languages since the early 1940s. The early breakthroughs were based on PSGs and more sophisticated grammar theories building on a basis of PSG rules (e.g., transformational grammars; Chomsky, 1963). Meanwhile, these theories have again been replaced in what is sometimes called the standard view in linguistics (Chomsky, 2000; Haegeman, 1991). Still, however, many modern syntax theories build on a kernel of PSG rules or a kernel of principles and parameters that translate into language-specific context-free grammars. When aiming at a neurobiological model of serial order, it may be advisable to consider this common basis of many theories of syntax.

What exactly is a phrase structure grammar implementing a context-free grammar or context-free rewriting system? As mentioned, it includes rewriting rules that are subject to one restriction: Only one symbol is allowed to the left of the arrow. At least one of these rules includes a *starting symbol*, abbreviated by an "S," to the left of its arrow. S can be used to start the generation process of a sentence. By applying a rewriting rule, the starting symbol is being replaced by other symbols. These are symbols for *syntactic* *categories*, which are sentence parts, such as subject, predicate or object, or noun phrase (NP) and verb phrase (VP). Additional rewriting rules can specify the replacement of syntactic categories by other syntactic categories, or the replacement of syntactic categories by *lexical categories*, which are word types sharing grammatical properties, such as noun (N), verb (V), article/determiner (Det), or preposition (Prep). Symbols for lexical categories finally can be replaced by symbols for actual words. Rules specifying the insertion of words are called *lexicon rules*. Thus, after replacing symbols by symbols and symbols by words, a word string or sentence results.

This can be illustrated by the following example rules:

(1) S → NP VP
(2) NP → (Det) N
(3) VP → V (NP)
(4) N → {Betty, machine}
(5) V → {laughs, cleans, switches}
(6) Det → {the, a}

Starting with Rule (1), S yields NP and VP, a noun phrase and a verb phrase. The noun phrase includes a noun, N, and can include an article or determiner, Det, as well. This is the meaning of Rule (2), in which the brackets to the right of the arrow indicate that the item within brackets, the determiner, is optional. According to Rule (3), the VP can expand to a verb, V, plus NP, but the NP is optional. The symbol for the lexical category N can be replaced by a real noun. Candidates are listed in Rule (4), which states that one of the words within curved brackets can replace the N symbol. Rules (4)–(6) are lexicon rules because they replace the abstract lexical category labels by lexical elements. In contrast, Rules (1)–(3) are syntactic rules in the narrow sense; they replace syntactic category symbols by other symbols for syntactic or lexical categories and are not concerned with individual words. By inserting (4) into (2) and (2) into (1), the string (7) can be derived.

(7) Betty VP

By additional insertions of (4) and (6) into (2) (yielding "the machine"), and of (2) and (5) into (3) (resulting in the VP "cleans the machine"), sentence (7) can be rewritten as (8).

(8) Betty cleans the machine.

Other correct grammatical sentences can also be generated based on the small set of rules. They include the sentences (9) and (10), but also strange word strings such as (11).

(9) Betty laughs.
(10) The machine cleans Betty.
(11) The Betty laughs machine.

More generally, a phrase structure grammar, as one possible description of a context-free grammar, includes

(i) symbols representing lexical categories (e.g., N, V),
(ii) symbols for syntactic categories (e.g., NP, VP),
(iii) rewriting rules with one syntactic or lexical category symbol to the left of the arrow (e.g., NP → Ar N), and
(iv) a starting symbol (S).

Phrase structure grammars specify sets of lexical and syntactic categories, or word and phrase types, that differ between languages. Modern syntactic descriptions attempted to replace such idiosyncrasies of individual languages by a more universal formulation of lexical categories, phrase types, and rules (Haegeman, 1991; Jackendoff, 1977). *Phrase types* are now defined with reference to their main lexical category, their head. Instead of verb and verb phrases, different projection levels of the verb are being distinguished. This makes the formulation of grammars of individual languages more economical. The equivalents of phrase structure rules can be derived from more abstract principles by setting a few language-specific parameters. As an example, the abstract principle might be that each word from a major lexical category X requires a complement (and forms a so-called X′ projection together with it), allows for an adjunct, and also requires a "specifier" to form a "maximal projection" (X″). The language-specific parameters to be set include, for example, the information about whether the specifier has to be placed before or after the verb. The syntactic categories would thus be considered derivatives, or *projections*, of one of the major lexical categories. Instead of VP, one would write V″, or V_{max}. This indicates that separate syntactic categories, independent of the major lexical categories they are driven by, are not necessary in a formal grammar [cf. (ii)]. The more systematic treatment of rules (on the basis of principles and parameters) and syntactic categories (considered to be projections of lexical categories) do not affect the general arguments made here.

At this point, it may be difficult to see why a network of neurons storing and processing syntactic information, called a *neuronal grammar*, should not be capable of solving the same tasks as a PSG does. One may speculate that a neuronal grammar could include

(i) neuronal elements representing morphemes and words
(ii) directed connections between neuronal representations

(iii) criteria for activity dynamics in the network and for string acceptance

and that these neuronal elements and dynamics could be as powerful in determining serial order of neuronal excitation as are formal grammars in determining the serial order of words and meaningful language elements. However, it has been argued that this is incorrect, that a neuronal device cannot, for principled reasons, process string types that can be handled easily by a formal model of syntax.

7.2 Center Embedding

The beauty of a context-free grammar lies in the fact that rules are recursively applicable. By repeatedly using rules, one can obtain syntactically correct sentences of considerable complexity. As pointed out by Chomsky (1963) and Levelt (1974), a neuron network of the McCulloch–Pitts type is equivalent to a *regular grammar* or *right-sided linear grammar*. A context-free grammar is more powerful than a regular grammar; it includes recursive rules that allow one to produce sentences in which other sentences are embedded, in which even further sentences can be contained. This "Russian doll" game can be played, as has been argued by linguists, with, in principle, no upper limit for the number of recursive operations, so that sentences of mild to extreme complexity can be synthesized. The fairly complex sentences would be characterized by *center embedding,* if the same sentence types are embedded in each other, by *self-embedding*.

Examples of center-embedded sentences or strings are listed as (12)–(14). Brackets have been inserted to help to structure the strings.

(12) Betty gives the machine (that had been cleaned the other day) to Bob.
(13) The rat (the cat (the dog chased) killed) ate the malt.
(14) Anyone ($_1$ who feels ($_2$ that ($_3$ if so many more students ($_4$ whom we haven't actually admitted$_4$) are sitting in on the course ($_4$ than ones we have$_4$) ($_4$ that the room had to be changed$_4$)$_3$) then probably auditors will have to be excluded$_2$)$_1$) is likely to agree that the curriculum needs revision.

Clearly, the two more extreme word strings, (13) and (14) (the latter being taken from Chomsky, 1963), are confusing and improbable. Sentence (13) is not easy to understand, and paper and pencil are necessary for working out a possible meaning for (14). Nevertheless, the argument maintained by some theoretical linguists would be that the internal grammar mechanism itself would, in principle, be capable of generating and processing such strings, were there not the theoretically not interesting limitations to memory and processing capacity. In fact, the "English sentence" (14) was used by Chomsky "to illustrate more fully the complexities that must in principle be accounted for by a real grammar of a natural language" (Chomsky, p. 286, 1963).

Example strings (12)–(14) show that there is, in fact, an upper limit for the language processing device in the human brain. The fact that much effort is needed, and even pencil and paper, to make sense out of (14) demonstrates that the brain-internal grammar mechanism is limited with regard to this analysis. There appears to be no empirical basis for a brain-internal grammar mechanism capable of dealing with strings such as Sentence (14) without external aids.

There are languages other than English in which center embeddings are more common. The German translation of (13) might be considered much less irritating by many native speakers of German than its translation, sentence (13), is by native speakers of English.

(15) Die Ratte [die die Katze (die der Hund jagte) getötet hat] ass den Malz auf.

Thus, there can be no doubt that center embeddings are possible, at least in certain languages. For a realistic model of language, the relevant question is what the upper limit for the number of embeddings is. Empirical research on this issue indicated that three embedded sentences are at the upper limit and are already difficult to understand (Bach, Brown, & Marslen-Wilson, 1986). This study also revealed that speakers of Dutch understand sentences including multiple crossed dependencies better than do speakers of German with center-embedded sentences. The authors argued that this rules out a pushdown mechanism (cf. p. 131f) as a universal basis of the human parsing mechanism. However, it may well be that a pushdown mechanism is available, but a distinct mechanism supporting crossed dependencies is available as well, and is more effective under conditions of high information load (Pulvermüller, 1994).

An abstract description of the problem posed by multiple center embeddings is as follows: In many languages there are strings such as (16) that later require complement strings in the reverse order, as schematized by (17). Here, the upper-case letters represent words, morphemes, or larger constituents including words. An initial string part would therefore require a complement string whose constituents exhibit the mirror-image sequence of the initial sentence part. If (16) is the initial part, (17) would be the mirror image part, and (18) would be the center-embedded construction.

(16) AB . . . M
(17) M′ . . . B′A′
(18) AB . . . MM′ . . . B′A′

Here, all pairs I, I′ are assumed to be complements of each other. If an element I has a complement I′, it means that if I occurs in a string, I′ must also be present. In (18), the sequence of constituents is therefore followed by the mirror-image sequence of their complements. Sentence (13) can be described by the mirror-image order of syntactic categories in (19).

(19) NP_1 NP_2 NP_3 VP_3 VP_2 VP_1

Here, each sentence consists of two complements, one NP and one VP. Numbers indicate the sentence to which the constituents belong.

To understand a center-embedded sentence such as (12), (13), or (15), the parts of each sentence that belong together must be related to each other. To complete sentence (13) after the third subject noun phrase has been uttered, not only information about the missing constituents of the main and subordinate clauses must be kept in memory, but information about their serial order as well.

Context-free grammars offer to deal with the phenomenon of multiple center embedding in the following way. A mirror-image string such as (20) can be generated by starting the generation process with the starting symbol, and by rewriting it as indicated in (21), then applying Rule (22), and finally Rule (23). (The optional "s" would be left out only in the latter rule application.)

(20) A B C C′ B′ A′
(21) s → A (s) A′
(22) s → B (s) B′
(23) s → C (s) C′
(24) B A C C′ A′ B′
(25) A A A A A′ A′ A′ A′

In these and similiar examples, lower-case letters, a, b, c, . . . , are used for syntactic category labels and upper-case letters, A, B, C, . . . , are symbols for actual parts of a string. Application of Rules (21)–(23) in a different order may result in (24), and repeated application of (21) in (25). Please note that the strings and Rules (20)–(25) are more abstract and simpler than the earlier examples taken from, or applicable to, real languages.

PSG Rules (21)–(23) would express that syntactic category symbols (indicated here by small letters) can be rewritten as strings of lexical category symbols (capital letters) and other, in this case optional (brackets), symbols for syntactic categories. After the arrow, the syntactic category symbols can appear to the left or right of a lexical category label. This is an important feature of context-free rewriting systems. If a syntactic category symbol appears only on one side of a syntactic category symbol, and if rules with more than one syntactic category label to the right of the arrow are not permitted, the descriptive power of the grammar is reduced. Such a one-sided linear grammar then, for example, is not able to process center-embedded strings (see Levelt, 1974). In right-sided linear grammar, the rules have their lexical category symbol always to the right of the lexical category symbol, and are therefore of the form (26).

(26) a → B c

Again, the small and large letters refer to syntactic categories and string parts, respectively. As mentioned, right-sided linear grammars are also called *regular grammars* and are equivalent to networks of McCulloch–Pitts neurons. Thus, it appears that the McCulloch–Pitts devices lack something context-free grammars exhibit.

The memory mechanism postulated to be necessary for the processing of a context-free grammar is called *pushdown memory*. Context-free grammars have been shown to be equivalent to *pushdown automata*, one type of linear-bounded automaton. In contrast to a finite automaton of the McCulloch–Pitts type, which is characterized by a finite number of neuron elements, each with a finite number of possible activity states, a pushdown automaton includes a pushdown store of unrestricted capacity. It is called *pushdown memory* because it is accessed in a first-in last-out manner, like a stack. New information is entered on top of the stack, above the older entries, and memory access is from top to bottom. Whatever piece of information is being entered into the store first is read out only after all information put into the store later has been removed. Conversely, the last piece of information entered into the store is read out first.

Having a pushdown store available makes it easy to store the knowledge about the mirror-image sequence of the verb phrase complements in example sentences (13) or (15), or the complements in the strings (19), (20), (24), and (25). Whenever an initial element is being perceived or produced, information about its complement, as specified by the respective rule [e.g., (21)–(23)] can be entered into the memory stack. The complement entered when the last constituent of the initial sentence part is being used remains on the top of the stack. Only after it has been read out is it possible to access the information about the complement of a constituent appearing earlier in the initial part of the string. This continues until the information about the complement of the initial constituent of the sentence has been read out. The way the pushdown automaton stores and accesses its information thus produces the mirror-image sequence of complements.

Both a regular and a context-free grammar, a McCulloch–Pitts device and a pushdown automaton, can generate, or specifically respond to, strings of arbitrary length. It is sometimes claimed that syntax algorithms are special, because they allow the generation or acception of, in principle, an infinite number of possible strings based on a finite set of rules. However, this is also true for regular grammars and their equivalent, McCulloch–Pitts neuron ensembles. This is explained in the context of neurons specifically responding to events of arbitrary length, the All and the Existence neuron in Section 6.1.

Another feature sometimes considered to be specific to syntax algorithms is the recursive applicability of their rules. However, this is, again, also a feature of regular grammars. An example of a simple abstract regular grammar, which is analogous to the context-free grammar (21)–(23), follows. Clearly, the rewriting Rules (27)–(29) also define a grammar, in this case a right-sided linear grammar, in which each rule can be applied recursively in the processing of a sting.

(27) s → A (s)
(28) s → B (s)
(29) s → C (s)

These rewriting rules can generate, for example, strings of A's, B's, or C's of any length, or a sequence in which these three symbols are mixed arbitrarily. The grammar is called *right linear* because, after the arrow, all syntactic category symbols (s) are to the right of lexical category symbols (capital letters). Analogs in terms of McCulloch–Pitts circuits are obvious. For example, a neuron activating itself can produce an unlimited sequence of activations of the same type. (if certain abstractions are being made).

One of the string types regular grammars cannot process are center-embedded strings of indefinite length. It has been argued by theoretical linguists that center-embedded strings are characteristic of many languages and that it is, therefore, important for grammar theories to allow for their processing without upper limit for the number of embeddings allowed. Note that this argument depends crucially on the assumption that extremely complex center-embedded strings such as (14) are relevant in grammatical processing. From this perspective, it appears that context-free grammars are superior to regular grammars with regard to the description of natural languages. Regular languages, and their equivalent neuronal automata, were therefore considered inadequate as descriptions of the syntax of natural languages. The neuron-based devices were abandoned by linguists, if they had at all been considered in the first place. A door between the language and brain sciences was closed, or possibly never opened.

This historical move is surprising because not too long after the elaboration of context-free PSGs and syntactic theories building on them, it could be shown that neuron-based models can be made equivalent to PSGs in describing context-free rewriting systems. The major difference between a McCulloch–Pitts network, with its finite number of possible states, and the pushdown automaton, with its stack of indefinite depth, obviously lies in their storage capacity, which is limited in one case but unlimited in the other. It is trivial that a device with unlimited storage capacity can outperform a device with limited capacity on certain tasks. Therefore, the obvious move was to think about possibilities to upgrade finite automata. For example, recursive transition networks were proposed and shown to be equivalent to PSGs. These formed the basis of a new class of grammar devices, called *augmented transition networks* (see also Section 6.1; Kaplan, 1972; Winograd, 1983; Woods, 1973). A different approach was rooted in the idea that McCulloch–Pitts networks could be allowed to grow so that the number of their neurons, and therefore the amount of information they can store, is unlimited (Petri, 1970; Schnelle, 1996b). It is unclear whether this view is realistic as a description of brain-internal processes. However, it is uncontroversial that a McCulloch–Pitts network with unlimited storage capacity and therefore a putatively infinite number of possible states is equivalent to a Turing machine, and is therefore even more powerful than a pushdown device (Schnelle, 1996b). It appears that the view about the inferiority of McCulloch–Pitts approaches to syntax cannot be maintained.

In summary, the limited vs. unlimited storage capacities of the respective devices is not an important difference between approaches to serial-order processing in the McCulloch–Pitts tradition on the one hand and those building on abstract pushdown automata on the other. Not only can this difference be compensated for easily by extending neuronal networks, its relevance is also questionable, given the dubious status of unacceptably complex strings with multiple embeddings such as (14). As mentioned, empirical evidence (Bach et al., 1986) suggests that speculations about unlimited pushdown stores are not required.

Perhaps a more important difference lies in the way pushdown automata and McCulloch–Pitts networks treat center-embedded sentences of finite length. Clearly, a McCulloch–Pitts automaton can also store information about a finite number of constituents that occur in the input. If a network is confronted with a string ABC, neurons each specifically activated by one of the string elements could store that the respective elements were present. The mechanism could be the same as those underlying the Existence neurons described in Section 6.1. However, although in this case information about the elements would be stored, the information about their serial order would be lost. One way to store the information about the sequence would be by way of sequence detectors – as also discussed in Section 6.1 – that respond only if input elements have been active in a particular order. However, for the three elements A, B, and C, there would, theoretically, already be 3^3 – that is, 27 possible sequences – each requiring its specific mirror-image complement. Postulating so many different sequence detectors may render the proposal unrealistic and therefore unattractive. It appears important to think about how a realistic neuronal device could store information about the important properties of such sequences of elements without the need for introducing new sequence detectors for each and every possible series of constituents. One approach is suggested by the pushdown storage mechanism: Perhaps each of the neuronal elements is active and some property of their activation retains the information about their serial order. This idea may appear attractive, because it allows for more economical information storage. However, it certainly needs elaboration, in particular with regard to possible realistic neuroscientific under pinnings. This is discussed in Chapter 12.

7.3 Discontinuous Constituents and Distributed Words

Is it possible that a neuron-based approach to syntax is more elegant for describing aspects of serial order in language than classical linguistic approaches rooted in the tradition of PSGs?

In Section 7.1, an example grammar was presented including six rewriting rules, (1)–(6) (see p. 126). Whereas these rules allow for the generation of several strings, for example, (9)–(11), they do not allow for generating sentence (30).

(30) Betty switches the machine on.

The problem here is that the last word's placement would not be specified by any of the earlier rules. At least one rule would need to be modified. For example, Rule (3), which is repeated here as (31) for convenience, could be replaced by (32):

(31) VP → V (NP)
(32) VP → V (NP) (on)

Given Rule (32) is being used, sentence (30) can be generated. The verb could then be followed not only by the noun phrase ("the machine") but by the verb particle ("on") as well. The downside of this formulation in (32) is that the intimate relationship between the two words *switch* and *on* disappears in this representation. The relationship between these two words is so close that the two can actually be considered one single composite word. This would in no way be a strong assumption. There are languages, for example, German, in which (33) is translated as a single word (34), which can also appear in its distributed form (35).

(33) switch . . . . . . on
(34) einschalten
(35) schalten . . . . . . ein

The proposal to call items such as (33)–(35) *distributed words* dates back to Harris (1945), a linguist. Distributed words or morphemes are a special case of *discontinuous constituent*, grammatically closely related sentence parts that are separated from each other in time in spoken language and separated in space in written language. Discontinuous constituents allow for, or can even require, that other material be inserted between the parts of the distributed representation.

The example of distributed words is even more interesting than that of discontinuous constituents at the level of larger syntactic categories. The main parts of a main clause, noun and verb phrases, separated from each other by a relative clause in center-embedded sentences, such as (12)–(15), would be an example of discontinuous constituents above the level of words. As discussed in Section 7.2, center embeddings can easily be dealt with by context-free rewriting systems. This is not so for distributed words such as (33) and (35). The reason is that the word is usually assumed to be dealt with as a whole. The grammar would process the word as a whole although the different parts occur at different places in the sentences and are separated in time or space. In a PSG, a lexicon rule would specify the two word parts and, in the generation process of a sentence, this lexicon rule would be applied only after the VP has been broken down into V and NP [Rule (3)]. Thus, one would ultimately end up with a string of words in which verb and verb particle are next to each other, although they may actually appear, for example, on opposite sides of an NP in the sentence.

A sentence including a discontinuous constituent or distributed word remains grammatical if the string inserted between its parts is made longer and longer. This is illustrated by examples (36)–(39), in which the two parts of the discontinuous constituent appear in italics. (The examples have already been mentioned in Chapter 6 and are repeated here for convenience.)

(36) Betty *switched* it *on.*
(37) Betty *switched* the coffee machine *on.*
(38) Betty *switched* the nice brown coffee machine *on.*

(39) Betty *switched* the coffee machine she received as a present from her mother at her 25th birthday *on*.

Admittedly, insertion of a longer phrase between the verb and the particle at the end of the sentences may sound somewhat odd. In English, one may, in these cases, prefer to place the particle immediately after its verb stem, as shown in (40).

(40) Betty *switched on* the nice brown coffee machine.

Nevertheless, there are again languages in which the constructions that are less usual in English are common or even preferred in language use. In German, sentence (39) would read as (41), in which, again, the two parts of the discontinuous constituent are in italic.

(41) Betty *schaltete* die Kaffeemaschine, die sie von ihrer Mutter zum 25. Geburtstag bekam, *ein*.

Discontinuous constituents can illustrate that, in language, sometimes very distant elements belong together, or in other words, are in a *dependency relationship*. As already mentioned in Chapter 6, this phenomenon is also called *long-distance dependency*. The word chain to be inserted between the two closely related constituents can be long. To produce or understand sentences such as (36) or (41), it is therefore necessary to be prepared to store information about the earlier of the mutually dependent elements, or the information that the second is required, and there is no apparent upper limit for the time this fact must be kept in memory.

Formulated in a more abstract manner, the problem of discontinuous words (assuming that there is a problem) is the following: Suppose the verb included in a discontinuous constituent can be part of strings of two kinds, either (42) or (43).

(42) A *switches* B C *on*
(43) A *switches* B C

Again, the letters represent (strings of) one or more morphemes. Related example sentences are (44) and (45):

(44) Betty *switches* the machine *on*.
(45) Betty *switches* the machine.

Now the problem is that the elements in italics are part of the same word, or, more generally, of the same distributed representation, although they can be separated in time by a long word chain.

To generate strings (42) and (43), the following rewriting rules (which form a one-sided linear grammar) may be used:

(46) a → A b
(47) b → switches c
(48) c → B d
(49) d → C (on)

Starting with a, the sequential application of Rules (46)–(49) yield string (42) or (43), depending on which of the two alternatives allowed by (49), with or without particle, is being chosen.

The parentheses in (49) suggest that the choice of the final particle is an optional process. However, this is not appropriate. Selection of the particle must be coupled to the selection of one of the two readings of the verb. Notably, the two verbs *switch* and *switch . . . on* in (42) and (43) have different meanings: (43), but not (42), implies that there are two machines, one of which is being replaced. There is therefore motivation to distinguish two different verbs that share the phonological form *switch*.

A tentative idea would be to represent the two parts of a distributed word in a neuronal grammar by using bistable elements, for example, the Existence cells discussed in Chapter 6. The two parts of a distributed word could be implemented by two Existence cells that are in some way functionally related. If one of the two is activated, the other one would also be aroused to some degree, and thereby prepared to accept the second part of the distributed word. The bistable nature of the Existence cells would guarantee that time lags of some length can intervene between the two complement word parts. This tentative idea would also need to be worked out in detail.

The problem of distributed words and other long-distance dependencies is virulent not only for one-sided linear grammars, but also for any context-free rewriting system. In order to solve it, different solutions have been proposed. For example, the two complementary constituents have been assumed to be generated together and one of them transported away from the other to its correct position by a movement or transformation (Harris, 1945, 1951, 1952, 1957, 1965). In sentence (44), the verb particle *on* would therefore be moved away from the verb root to its sentence final position. The idea that the grammatically correct sentences of a natural language are the result of movements of constituents within strings first "base-generated" by a context-free rewriting system has been worked out in great detail and in various forms by Chomsky (1957, 1965).

A different solution of the problem of discontinuous constituents requires introduction of new categories of nonterminal nodes. One example is the *slash features* in modified PSGs (Gazdar, Klein, Pullum, & Sag, 1985; Joshi

& Levy, 1982) that specify syntactic objects lacking a complement or are otherwise incomplete. According to this proposal, the information about a missing verb particle could be, so to speak, transported through the sentence representation. The nodes of the syntactic tree intervening between the two complements (e.g., verb root and particle) would incorporate the information that a particular type of complement (e.g., the particle) is missing.

Still a different solution of the problem posed by discontinuous constituents and long-distance dependencies is offered by a more fine-grained categorization of words into lexical subcategories, as first proposed in the context of valence theory and dependency grammar. Assume we wish to represent the root of the particle verb and the verb without particle as two distinct elements in the lexicon, say switch_1 (root of transitive verb requiring the particle, abbreviated here as "V14par") and switch_2 ("plain" transitive verb, "V14"). The different meanings of the word *switches* may provide sufficient motivation for introducing two word representations in this case. (The idea of more fine-grained lexical subcategorization also underlies the dependency rules formulated in the next section.) This would have the consequence that the rewriting system would need to become more complex than in (46)–(49). These rules and the syntactic categories they operate on must be doubled to distinguish between the representations triggered by the two different verbs.

(50) b → switches_1 c1
(51) c1 → B d1
(52) d1 → C on

(53) b → switches_2 c2
(54) c2 → B d2
(55) d2 → C

Technically speaking, it is now guaranteed that, if switches_1 is chosen by application of Rule (50), the particle *on* must follow at the end of the sentence. However, choice of the other variant of the verb by Rule (53) Rules out a final verb particle *on*. This makes the syntactic description more precise at the cost of introducing more material, lexical categories, and rules, into the right-linear (regular) grammar, in this case.

However, there is still nothing in the one-sided linear grammar that would allow for representing the fact that it is precisely the verb stem and the particle that are interdependent. This disadvantage could be overcome, for example, by using one of the strategies outlined previously (making use of movement or slash features) or by a neuronal machinery implementing the close relationship between complements by direct links between their representations and bridging the time delay by bistable neuronal elements (i.e., Existence neurons, memory cells).

Other disadvantages of the more complex right-sided linear grammar introduced for solving the problem of discontinuous constituents and distributed words are that so many different rules are now necessary and also that the number of symbols for syntactic categories was doubled. If there were several discontinuous constituents or long-distance dependencies in a sentence, the number of additional rules and nonterminals would multiply. If the number of long-distance dependencies is n, the number of nonterminals and the number of rewriting rules would increase with 2^n. This exponential increase is not attractive. In a neuronal circuit in which discontinuous constituents are realized as linked bistable units, the number n of elements to be kept in memory can be represented by n memory circuits.

This example may indicate that, although finite neuronal automata of the McCulloch–Pitts type and one-sided linear grammars are equivalent, there may be syntactic phenomena aspects of which can be coded more easily, or more elegantly, in a neuron network than in a rewriting system.

7.4 Defining Word Categories in Terms of Complements: Dependency Syntax

In PSGs, rules were proposed to operate on syntactic categories. A different approach is to base grammar algorithms on lexical categories (or word classes) and their complements, the necessary additions they require, and adjuncts, the optional additions they allow for. This alternative approach was taken in the 1950s by a syntactic theory, or class of theories, called *dependency grammar* (Tesnière, 1953, 1959).

Dependency grammars can be formulated in terms of dependency rules that operate on lexical categories and do not necessitate syntactic categories. These rules specify that the presence of an element of lexical category A in the string requires the presence of additional elements from lexical categories B, C, D, ... The latter elements are therefore said to be dependent on A.

To distinguish rewriting and dependency rules from each other in the present context, a different notation for the latter is used. Hays (1964) and Gaifman (1965) proposed to abbreviate the valence and complement structure of verbs by brackets in which symbols for the required complements are inserted either to the left or right of an asterisk symbolizing the position of the element outside brackets they are dependent from (for a more elaborate formalism, see Heringer, 1996). Here, a similar notation is used with an asterisk between slashes representing the position of the parent category and the labels of the dependent categories between brackets. Optional rules, that is, rules that specify possible but not required adjuncts, are shown in double brackets. Curly brackets frame alternative words that can replace labels for lexical categories. The context-free grammar (1)–(6) has its analog in the dependency grammar (56)–(65):

(56) V14par (N1 /*/ N4 Par)
(57) V14 (N1 /*/ N4)
(58) V1 (N1 /*/)
(59) N1 ((Det /*/))
(60) N4 ((Det /*/))
(61) V14par {switches}
(62) V14 {cleans}
(63) V1 {laughs}
(64) N {Betty, machine}
(65) Par {on}

This example of dependency grammar includes five syntactic dependency rules, (56)–(60), and five lexicon rules, (61)–(65). This formulation is slightly more detailed than the formulation chosen in (1)–(6) because it includes subcategories of verbs and nouns. In the present grammar fragment, nouns are either subjects, and are therefore in the nominative case, N1, or they are accusative objects, N4, and are therefore in the accusative case. The verbs are intransitive "one-place" verbs, V1, which do not require complements except for their subject noun, or they are subclassified as V14, transitive verbs requiring an additional accusative object, or, as a third possibility, as transitive particle verbs V14par, which require a verb particle in addition to an N1 and N4. The lexicon rules also become more detailed, consistent with the larger number of subcategories. Rule (64) applies for both N1 and N4. The increase in sophistication of the grammar and the larger number of rules now allow for a more precise adjustment of the grammar to the set of strings the grammar should generate.

Here are illustrations of the meaning of the dependency rules: (56) would read "a transitive particle verb stem requires a noun in the nominative case that precedes it and a noun in the accusative case and a particle that follow it." Rules (59) and (60) state that nouns require determiners that precede them.

The dependency system (56)–(65) allows for the generation of some of the example sentences discussed previously and are repeated here for convenience. Following each sentence, its structure is illustrated on the basis of the syntactic dependency rules. The dependency rule appropriate for the respective verb type is first selected and other rules are inserted to specify additional relevant lexical categories.

(66) Betty cleans the machine.
V14 (N1 /∗/ N4 (Det /∗/)) by insertion of (60) into (57)
(67) Betty laughs.
V1 (N1 /∗/) by insertion of (59) into (58)
(68) The machine cleans Betty.
V14 (N1 (Det /∗/) /∗/ N4) by insertion of (59) into (57)
(69) Betty switches the machine on.
V14par (N1 /∗/ N4 (Det /∗/) Par) by insertion of (60) into (56)

Note that sentence (69) and the respective syntactic structure representation including a distributed word can be represented. A particular dependency rule, (56), can express that there is a direct relationship between the verb stem and its particle.

Again, a main idea underlying dependency grammar is that words as members of a particular lexical category are characterized by, or, actually can be defined by the set of complement lexical categories they require. As mentioned in Section 7.3, this idea has spread to other approaches to syntax and has been incorporated in different ways into the respective algorithms. The grammar circuits offered in Chapters 10–12 draw heavily on this idea.

7.5 Syntactic Trees

The application of rule systems such as the ones exemplified previously can be illustrated using so-called tree graphs. Unlike real trees, these graphs have a root pointing upward and branches hanging downward. Trees are two-dimensional graphs that represent the application of rules schematically in the generation or analysis of a sentence. If a rule is applied, the symbol to the left of the arrow of rewriting rules or the symbol outside brackets in dependency rules – the *parent category* – appears above the symbols specified otherwise by a rule – called *child category* labels here. Parent labels are connected to those of their children by lines, as specified by rules. These lines are called *projection lines*. In the tree representation, the word string specified by a set of rules is given at the bottom, as the final leaves of the tree. The words and morphemes at the bottom of the syntactic tree are in the linear order in which they appear in the sentence. Trees are strictly two dimensional in the following sense: Projection lines are not allowed to cross. This property is sometimes called *projectivity* (Marcus, 1965). In many analyses, the two parts of a distributed word violate the projectivity principle.

Figure 7.1 shows how different graphs illustrate the structure of the English sentence (70).

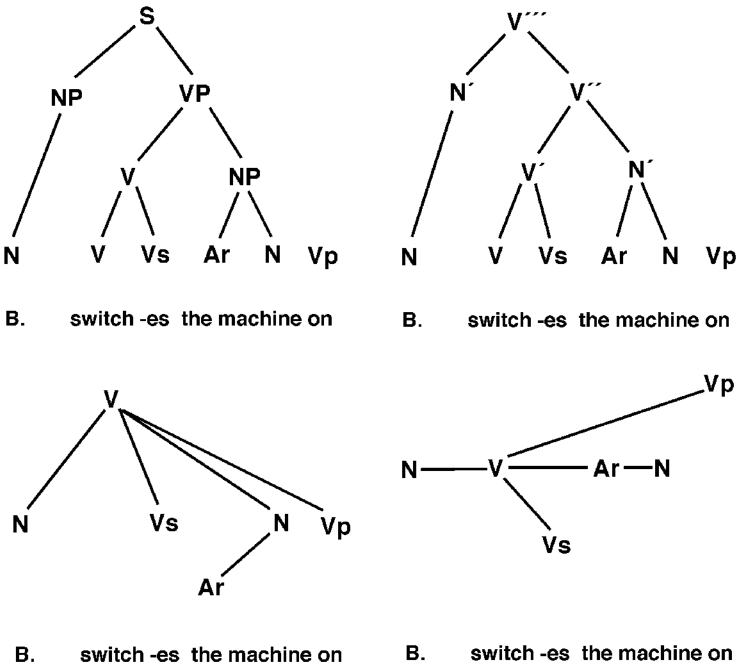

Figure 7.1. The structure of a sentence is illustrated based on different approaches to grammar. (Upper left) Context-free phrase structure grammar is used to analyze the string. (Upper right) X-bar syntax. (Lower left) Dependency grammar. (Lower right) A symbolism thought to represent neuronal elements and connections between them. (For explanation, see text.)

(70) Betty switches the machine on.

As mentioned, analysis of this string is not a simple task. Obviously, one problem is that it includes a particle verb that is realized as two separate constituents that belong together, *switches* and *on*. Note that in the trees, abbreviations of lexical and syntactic categories have been changed slightly to make them even shorter, and Vp replaces Vpar and Ar is used instead of Det. Also, the verb suffix is analyzed (Vs).

The diagram on the upper left in Figure 7.1 gives an example of a syntactic tree constructed from a simple context-free grammar or PSG. The macroscopic sentence structure is analyzed according to Rules (1) and (3) (Section 7.1), and the noun phase is broken down as detailed in Rule (2). In addition, the verb is rewritten as verb followed by a verb suffix. The analysis leaves open how the verb particle Vp should be linked into the construction. As mentioned in Section 7.3, one possibility would be an indirect link to the verb phrase so that a line connects the VP label to the Vp(ar). As mentioned, this would obscure the relationship between the verb root and its particle and would therefore not be optimal. The verb and its particle can be linked to each other directly only through a line if projection lines cross. This is, however, not allowed in this kind of graph, which is, as we know, conceptualized as strictly two dimensional and therefore without crossings of projection lines. A third possibility is to attach the verb suffix to the noun or noun phrase nodes directly and define the intervening constituents, the VP and NP (and N), as incomplete constituents missing a verb particle. In this case, the feature that the particle is missing could originate from the verb stem representation and be transported through the intervening nodes of the tree, finally reaching the constituent directly connected to the particle representation.

The tree graph on the upper right of Figure 7.1 is congruent to the PSG tree on the upper right. The only difference is the relabeled nodes. Each syntactic category label has been replaced by the name of a lexical category plus prime signs indicating the distance between the corresponding nodes. This is similar to an analysis based on X-bar theory. The charming thing here is that it becomes apparent that the syntactic categories are closely related to lexical categories. The problem with the attachment of the verb particle remains.

The diagram on the lower left of Figure 7.1 shows how the same string may be represented by a dependency grammar lacking nonlexical syntactic categories entirely. The tree is near equivalent with the bracketed representation in (69). The lines in this tree have a different meaning than in the two earlier graphs. Apart from relationships between lexical entries and lexical categories, they express the dependency relationship between lexical categories. In this formulation, it is possible to attach the verb particle to the verb it belongs to without violating the projectivity principle. This is not to say, however, that the projectivity problem is absent in this grammar type for other discontinuous constituents or unbounded dependencies (see Heringer, 1996). A major difference distinguishing the dependency approach from rewriting systems is that in the former the syntactic properties of one of the words in the sentence, in this case the verb, determines the structure of the sentence representation.

An ensemble of connected neuronal elements that may serve similar purposes as the aforementioned trees is sketched on the lower right of Figure 7.1. It is based on the idea that in the brain, there are specific representations of lexical categories. In this case, symbols for neuronal connections replace the lines indicating rewriting or dependency relations in the tree graphs.

There appear to be no *a priori* reasons for preferring one of these representations, although it appears at first glance that treatment of the verb particle is least problematic in the representations at the bottom. However, again, it is well known that after introduction of additional principles into

PSG-based rewriting systems, aspects of the problem of discontinuous constituents and unbounded dependencies can be solved (Gazdar et al., 1985).

Furthermore, note that none of the representations in Figure 7.1 captures the fact that there is agreement between the subject noun *Betty* and the verb suffix -*es* – the phonological realization of the present third person singular feature of the verb. For the graphs to describe such fine-grained structural relationships, some refinement would be necessary. A possibility is again offered by transporting features through the tree representation so that the feature can be shared by lexical elements distant from each other in the string. In the case of the agreement between subject noun and verb suffix, the feature “third person present singular” could climb through the tree, from the noun in subject position to the verb. Clearly, this would require a mechanism different from the one expressed in terms of rewriting or dependency rules or related descriptions of serial order in sentences. One may ask whether such an additional feature transport mechanism is required (Pulvermüller, 2002).

7.6 Questions for a Neuronal Grammar

Of course, there are many open questions that must be addressed by a neurobiological approach to serial order, such as

(a) How can center-embedded strings be represented (e.g., those of the form ABCC′B′A′)?
(b) How can discontinuous constituents and distributed words be realized (e.g., *switch ... on*)?
(c) How is it possible to specify the syntactic relatedness of distant elements in a string (e.g., noun and verb agreement)?

These problems have been dealt with in great detail in syntactic theories, but a brain perspective on them has not been offered, or at least not been specified in great depth.

More questions pose problems to a neuronal device designed to process serial order, one of which is the possibility to store the fact that the same lexical entry or syntactic category ocurs more than once in a sentence. Clearly, one word can occur repeatedly in a sentence, sometimes even with the same meaning.

(71) *The* dog chased *the* cat.
(72) *Die* Ratte [*die die* Katze (*die die* Hunde jagten) getötet hat] ass den Malz auf. (Engl.: The rat [whom the cat (whom the dogs chased) killed] ate the malt.)

In sentence (71), the determiner *the* occurs twice, whereas in example (72), the German word *die* is used five times – twice as a relative pronoun (i.e., *that* and *who*), and three times as definite articles marked for casus, numerus, and genus. It is clear that repeated use of words in single sentences is a rather common phenomenon, and that words are frequently used in the same grammatical function and semantic meaning when being repeated. The problem that a representation is being processed twice also occurs if a syntactic category is used repeatedly in a derivation.

Context-free rewriting systems offer to deal with the phenomenon, as discussed in detail in the context of examples (20)–(29). Rules are usually defined as being recursive, and there is no restriction on the number of occurrences of syntactic categories, lexical categories, or lexical items in one tree. In contrast, however, there are models for which an account of this feature is difficult. Localist connectionist models of language, for example, implement words as nodes in a network. If two occurrences of the same word in one sentence are to be modeled, it may be necessary to activate the representation twice simultaneously, so to speak, a possibility not offered by most models. Nevertheless, repeated access to the same representations – words, lexical or syntactic category representations, and/or rules – and storage of this multiple use is necessary for modelling important properties of language (Sougné, 1998).

This type-token problem is not easy to realize within a neuronal machine. Section 6.1 discusses one form of repeated processing of the same information when modeling a device responding to a string of several identical stimuli or events. However, this solution has the disadvantage of not keeping track of how many times the neural representation has been activated. It is clear that to produce or understand sentences (71) and (72) successfully, one must know not only that a string of several *the*'s or *die*'s was present, but also exactly how many of *the*'s or *die*'s were present to answer the question of whether this string was grammatical and acceptable; the same knowledge is necessary for comprehending the string successfully.

To realize repeated use of the same word or lexical category, one may postulate several neuronal representations of such an entity. For example, predicates have been proposed to be available a finite number of times in neuronal automata (see Shastri & Ajjanagadde, 1993). Processing of several tokens of the same type would then correspond to activation of two, three, or more neuronal devices. One may assume to have, for instance, five McCulloch–Pitts neurons representing the word *the*. Although this assumption is certainly possible, it is not very attractive because (i) it is difficult to explain why the brain should invent several representations of the same entity, and (ii) the choice of any upper limit for the number of representations would be arbitrary.

A second possible view posits that different levels of activity of one neuronal device could code for multiple processing of the same linguistic element. This solution, however, requires that activity levels coding 1, 2, and more tokens of the same type are distinct. Given the gradual loss of activity over time seen in many memory cells (see Chapter 2), this does not appear to be a fruitful strategy. Information about one or several occurrences of the same element may easily be mixed up if neuronal representations exhibit exponential activity decrease over time. There is cross-talk and, therefore, loss of information.

As a third perspective on solving the problem of multiple use of one linguistic entity, one may assume that multiple activity patterns in the same neuronal device realize presence of several tokens of the same type in a string or in the linguistic representation of a string. This possibility is explored further in Chapter 12.

Another question posing problems for a serial order network is related to the organization of lexical categories. All syntactic theories use lexical categories, and there is little doubt that some kind of a category representation must be postulated in a neuronal architecture of serial order as well. However, how this could be done is not clear. In some of the proposed neural networks, individual neurons or neuron ensembles are assumed to represent and process individual words or morphemes (see Chapters 4–6). For a network underlying the production and perception of syntactic strings, it is, for obvious reasons, desirable not to have separate representations for each chain of words; otherwise each word string would have to be learned separately, and the ability to generate new grammatical sentences could not be explained. Therefore, it appears advantageous to have a representation of what different words and morphemes have in common. If there is a neuronal representation of these common features of the words and morphemes of a particular type, rules of serial order could be realized as the connections not between individual word/morpheme representations, but as the connections between the representations of common features of several words (Pulvermüller, 2000).

To sum up, in addition to the problems (a) to (c), questions (d) and (e) must be addressed:

(d) How can repeated use of the same word or lexical category within a sentence be modeled and stored?
(e) How can lexical categories be realized in a neuronal network?

Tentative answers to questions (a)–(e) are proposed in Chapters 9–12.

CHAPTER EIGHT

Synfire Chains as the Basis of Serial Order in the Brain

The putative brain mechanisms of word processing are highlighted in earlier chapters; however, mechanisms of serial order have until now been addressed only sporadically in terms of neurons. This chapter now focuses on neuroscientific questions about serial order. Serial-order mechanisms relevant for language processing are being specified in great detail by linguistic theories, but these are not easily converted into neuron mechanisms. Such conversion of linguistic theory into neuroscientific models is possible by replacing rules and the operations they determine by artificial neurons and their dynamics (see Schnelle, 1996b). However, this chapter and Chapter 9 take a different approach. First, the question of which serial-order mechanisms are suggested by neurophysiological research is considered. The mechanisms inferred from neuroscientific data are then used as a basis of a neuronal model of syntax.

How may rules governing serial order of language elements be realized in the brain? This question may be asked with regard to the level of words and morphemes, the smallest language units that carry meaning, the meaning atoms, so to speak. A similar question may also be asked at the level of phonemes, language sounds distinguishing between morphemes or words: How are phoneme sequences stored and processed neuronally? The following paragraphs review brain mechanisms that could be relevant for establishing serial order of language units.

8.1 Neurophysiological Evidence and Neuronal Models

A sequence AB of events can be represented by two directly connected neuronal units: one corresponding to event A and the other to event B. If the respective neuronal units are referred to by Greek letters, α and β, the sequence of events could be realized as α, β, and, in addition, a direct connection from α to β. A single neuron α could, therefore, by way of its direct projection to neuron β, arouse it whenever active.

However, it is unlikely that single cortical neurons connected in this way play a role in language processing. As mentioned in Chapter 2, the connections of most neurons in the cortex are known to be weak so that input from one single neuron would usually not be sufficient to strongly enhance the firing probability of a second neuron on which the first projects (Abeles, 1991). Data from both *in vivo* and *in vitro* intracellular recordings show that individual postsynaptic potentials are small and unreliable (e.g., about $0.2\ \mu V +/- 0.1\ \mu V$ in the neocortex of the rat). Thus, the concept of strong connections between individual neurons is not realistic. Near-synchronous input through many of its afferent synapses is usually necessary to arouse a neuron. It appears therefore to be more likely that sets of neurons project onto each other, thereby making up broad neuron chains that determine spatiotemporal patterns of activity. Waves of synchronous excitation could then spread along these broad neuron chains.

Studies on the correlation of the firing in multiple neuronal units provided strong evidence for complex spatiotemporal activity patterns in which many neurons participate (Abeles et al., 1993; Vaadia et al., 1995). The firing probability of a single neuron could best be determined when more than one preceding neuronal event and behavioral context were taken into account. This context dependence cannot be modeled based on a chain of single neurons, each projecting onto the next in the chain. If this view were correct, one would expect each neuron to have a constant influence on the firing probability of other neurons. Instead, it was found that two neurons firing with a particular exact delay were good predictors of the firing of a third neuron. In contrast, each of the former neurons firing alone or both neurons firing with a slightly different delay would not have a detectable effect of the third one. Thus, the firing of a neuron was found to be strongly dependent on the context of more than one other neuronal events.

The complex context dependence of firing probabilities follows from a model in which groups of neurons are connected in chains. In this case, the synchronous activity of one of the groups that are connected in sequence is necessary to arouse the next set. This type of neuronal circuit has been labeled a *synfire chain* (Abeles, 1991). Figure 8.1 illustrates a simple model of a synfire chain. In this illustration, each neuron needs two simultaneous inputs to become active, and each of the sequentially connected sets of the chain includes three neurons. This is a simplification made for ease of exhibition; the number of neurons of each neuron set connected in sequence is probably higher, between 50 and 100 neurons (Diesmann, Gewaltig, & Aertsen, 1999), and their firing threshold is probably in the order of 5–10

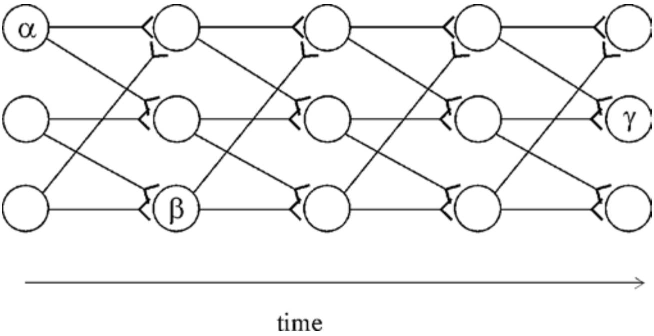

Figure 8.1. A synfire chain consisting of five subgroups of neurons connected in sequence. Circles represent neurons and v-shapes connections between cells. Each neuron is assumed to have a threshold of 2. Each subgroup activates only the next subgroup if all of its neurons fire simultaneously. If the synfire chain is active, the neurons α, β, and γ embedded in the chain will fire in a fixed temporal order (e.g., β 10 ms after α, and γ another 30 ms after β). If α and β fire with this fixed delay, 10 ms, the probability for γ to fire another 30 ms later is enhanced. If α or β fire alone, and therefore independently of the synfire chain, then the firing probability of γ is not affected.

simultaneous inputs (Abeles, 1991). In the model circuit, simultaneous activation of the three neurons on the left leads to a wave of activity spreading to the right. The serially connected subsets, or *steps*, of the synfire chain activate each other after exact delays. The neurons of each subset activate the neurons of the next subset at exactly the same point in time.

Figure 8.1 explains the relationship between the assumed synfire mechanism and the mentioned data on context dependence of neuronal firing. Note that neurons α and β in this example circuit can be active independently of each other. Given the signals they send out are weak, each of them being active alone would have only a negligible influence on neuron γ at the end of the chain. However, if they are active in a given temporal order, for instance β 10 ms after α, this would likely be because the chain as a whole is active, in which case γ would be activated at a fixed delay after β, perhaps 30 ms later. Thus, the firing probabilities depend strongly on the context of other neuronal firings. The conditional probability of γ firing 40 ms after α and 30 ms after β is high, but the firing probabilities of γ in other contexts for example, after α has fired but not β, may be low.

The synfire model implies that a cortical neuron can be part of several synfire chains and that it can therefore be a member of different spatiotemporal firing patterns. The individual neurons would then become active, on a regular basis, in different well-defined behavioral and neuronal contexts. In multi-unit studies of cortical activity, it was found that much of a neuron's activity that would otherwise be considered to be "noise" or "spontaneous activity" could actually be described in terms of frequently occurring spatiotemporal firing patterns (Abeles et al., 1993, 1995).

To further illustrate the synfire mechanism, a schematic representation of two intersecting synfire chains is shown in Figure 8.2. Whenever the neuron group at the upper left is active, an excitation wave spreads downward, terminating at the lower right. Although the neurons in the very center are also heavily connected to the neuron groups at the lower left, activity on the lower left dies out in this case. In the same way, a wave from the upper right spreads only to the lower left. There are two distinct spatiotemporal patterns of activity that are prevented from getting mixed up by the very nature of the connections, although the structural bases for these patterns overlap. Two of four of the neurons in the central layer, where the two synfire chains cross and the two ovals overlap, belong to both synfre chains. Each of the other neurons included in the synfire chains may also be part of other chains as well.

An essential feature of the synfire model is that information highways share part of their structural basis, which has the obvious consequence that each neuron's firing depends on the firing context of several other neurons. As mentioned, the shared part of the two chains sketched in Figure 8.2 are the neurons in the middle of the central neuron group, where the ovals intersect. These neurons would become active as part of an activity wave

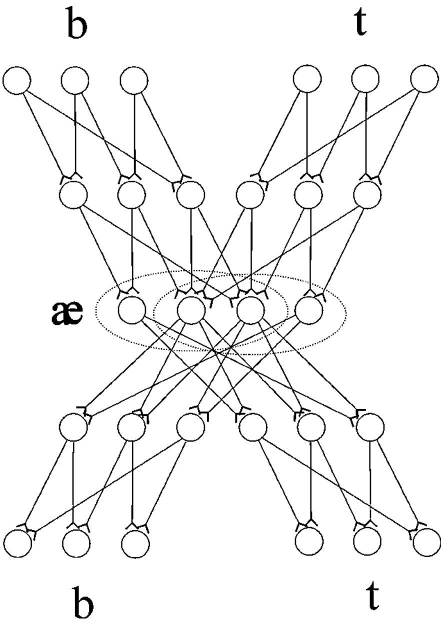

Figure 8.2. Synfire chains that cross. Each circle represents a neuron and y-shapes represent connections between neurons. Each neuron is assumed to have a threshold of 2. Possible phonemic correlates of subsets of the synfire chains are indicated.

starting at the upper left, but would also be activated if an activity wave started at the upper right. In this case, the firing of these neurons shared by the two synfire chains alone would not decide the direction in which the excitation wave progresses. However, given these *shared neurons* are active, the direction of activity flow can be determined if more of the context of the firing of these neurons is taken into account. The left- and right-most neurons in the central group, which are included in only one of the ovals, have the role of *context indicators* channeling the wave of activity either to the left or right. Synfire chains can thus have shared neurons, in which case context indicators are crucial for keeping them and the spatiotemporal activity pattern they determine separate.

8.2 A Putative Basis of Phonological Processing

Synfire chains have been proposed as a putative neuronal basis of articulatory programs (Braitenberg & Pulvermüller, 1992). The exact timing of nerve cell firings determined by the circuitry would allow for realizing precisely timed articulations. From a cognitive perspective, the beauty of the synfire chain mechanism lies in its potential to provide a straightforward solution to what Lashley (1951) described as one of the main aspects of the *problem of serial order in behavior*. If each phoneme or letter was represented as a separate entity, the possible words of a language could not be modeled simply by direct connections of 50 or so phoneme or letter representations. Too many different sequences would be possible whenever a set of phoneme or letter representations are selected. If a set of representations is activated, for example, the phonemes [t], [æ] and [b], there would be no information about their serial order, so that different sequences would be permitted (e.g., "tab" and "bat"). However, if not phonemes but context-sensitive phoneme variants were represented instead, each possible sequence could be realized by direct links between neuron sets. The context-sensitive variant of phonemes would be, for example, [b] at word onset followed by [æ] – which can be abbreviated #Bæ – [æ] following [b], and followed by [t], bÆt, and [t] terminating a word and preceded by [æ], æT#. The three context-sensitive phoneme-like units #Bæ, bÆt, and æT# would determine the elements of the phoneme sequence and their serial order. A similar solution to the serial-order problem has earlier been suggested by Wickelgren (1969).

In contrast to Wickelgren's approach, which acknowledges only phoneme-context effects among adjacent phonemes, it is well known that the physical properties of the realization of language sounds can depend on more distant context. The synfire model would suggest that for each frequently encountered word, a synfire chain is being established and that the synfire chains of different word forms intersect, cross, and overlap depending on how the *physical features of the sounds* are being shared between the usual realizations of the word forms.

In phonology and phonetics, context-dependent variants of phonemes are sometimes called *allophones*. Allophones are phonotactically determined variants of a language sound with complementary distributions, the variation being determined by their position relative to other phonemes in a string. There are also *free variants of phonemes* without obvious phonotactic determinants and similar distributions (e.g., the [r] and [R] sounds generated by the tip vs. back of the tongue as they coexist in German and French). It is possible that the context-sensitive variants of a phoneme that are laid down in the synfire chain include phonetic features of both phonotactic phoneme variants and of the free phoneme variants the individual happened to be frequently exposed to. The synfire model allows for specifying a putative underlying mechanism of the processing of language sound variants.

It may be worthwhile to illustrate the processing of phonological and phonetic information in a synfire architecture. Figure 8.2 can be used for this purpose, although the relevant articulatory–phonological mechanisms are much simplified. The focus is on articulatory mechanisms only, although analogous mechanisms must be postulated on the acoustic side.

If the synfire chain starting at the upper left and running to the lower right is considered the neuronal correlate of the syllable [bæt], its component neuron groups can be taken as the correlate of the relevant linguistic elements, the language sounds as they are realized in the respective context. Each sound representation would be composed of two different kinds of neuronal elements: One related to invariant properties of the articulation of a phoneme and realized by the shared neurons of different chains; the other related to variations of the articulation depending on the context sounds in which a given language sound is being articulated, realized by the context-indicator neurons mentioned previously. These would cover features of allophones, the context variants of phonemes.

This can again be made more plastic on the basis of Figure 8.2. The neurons shared between the two representations of context-sensitive variants of the phoneme [æ] – the two middle neurons in the central layer of Figure 8.2 shared by both chains – could relate to articulatory distinctive features of that phoneme (e.g., lips open but not rounded, tongue at the bottom of the mouth). In contrast, the neurons deciding between the possible successors and distinguishing between the alternative synfire chains – the left- and rightmost neurons of the central neuron group, the context indicators – would process information about the articulatory movements determined by the context, the preceding and following articulations. Information about differences between allophones is stored and processed here.

In this view, the neurobiological equivalent of a phoneme in context would therefore consist of neurons related to *articulatory distinctive features* (the shared neurons) and others realizing *context-sensitive phoneme variants* allophones (the context indicators).

Any discussion of a possible neurobiological correlate of phonological processing must include some limitations of the proposal. Clearly, the synfire model postulates exact delays. Although this is an advantage for modeling precisely timed articulations, the model circuits may provide not enough flexibility for explaining certain obvious phenomena. For example, speech rate can be high or low. If synfire chains were designed to model different articulation speeds (i.e., rapid or slow pronunciation of a syllable), it would be necessary to allow them to be tuned so that the basic time scale on which they operate can be changed. One possibility is to assume that the chains operate at different levels of background activity affecting all neurons in the chain to a similar degree. Weak background activity slows the speed with which activity runs down the chain. The candidate explanatory process is that more time is needed for temporal summation of neuronal activity necessary for activating each step in the chain. Higher background levels of activity may speed up the chain because less temporal summation would be needed. It is clear that this proposal must be worked out in more detail, but it is equally clear that such elaboration is possible (Wennekers & Palm, 1996).

In assuming that synfire chains realize the articulation of phoneme strings, one may suggest that each of the functional webs underlying word processing (see Chapter 4) could include a synfire chain controlling the articulation of the word form. The functional web representing the phonological word form would include, or would have attached to itself, a chain of neurons whose spatiotemporal sequence of activity is timed precisely. Ignition of the functional web could therefore include or be followed by the sequence of neuron firings determined by the articulatory synfire chain. Ignition of the word-related functional web would not necessarily also activate the synfire chain. Such activation of the articulatory program may require that the respective neurons be in a preparatory state.

This proposal does not address the fact that words can be realized in different ways, some of which are substantially reduced in rapid speech. For example, the function word *will* (but not its noun homophone) can be reduced to *ll*. To model this in a network, one may attach two or more mutually exclusive synfire chains to the representation of the function word. As an alternative, one may model the reduced articulation on the basis of background activity levels. If articulation speed is very high, the articulatory network may be under strong background activation, so that the respective chains become too fast for the neuromuscular articulatory machinery to follow their commands. This results in omissions of sounds, in particular in contexts in which fast complex articulatory movements would be required (e.g., "it comes" realized without the [t]).

Keeping in mind the limitations of the proposal, it still appears that the synfire model may allow for a fruitful approach to phonological brain processes. Because it offers (i) a simple solution for one aspect of the serial order problem, (ii) a mechanism for precisely timed articulations, and (iii) a mechanism for coarticulation effects, the synfire model may provide a neurobiological perspective on articulatory mechanisms. Clearly, this does not prove the model correct. There are also fruitful approaches to memory processes for phonological sequences that exploit alternative, or possibly complementary, mechanisms (Page & Norris, 1998).

One of the features this proposal shares with recent psycholinguistic approaches (Marslen-Wilson & Warren, 1994) is that it does not require separate representations of phonemes. Overlapping sets of neurons related to phonetic distinctive features and context features are proposed to correspond to the contextual variants of each language sound. The intersection, or, as an alternative, the union of these overlapping neuron sets, can be considered the (putative) neurobiological basis of a phoneme. No additional "phoneme nodes" are required; rather, several overlapping context-sensitive representations of a phoneme coexist. Furthermore, feature overlap of all context vatiants of one phoneme are not required. The family resemblance argument (Wittgenstein, 1953) may hold for phoneme variants as it holds in the semantic domain (Section 5.2.3), and the underlying neuronal mechanisms may be similar. Although the intersection of context-sensitive representations underspecifies the articulatory movements necessary for realizing the phoneme, each of the context-sensitive representations itself overspecifies the articulatory movements of the phoneme because it also includes redundant information about phoneme variants whose features may, strictly speaking, not *all* be necessary for distinguishing word meanings.

8.3 Can Synfire Chains Realize Grammar?

It is tempting to apply the synfire model to higher-order sequences of meaningful units, morphemes, and words. One may want to define and neuronally implement a word's syntactic role on the basis of its *context words*, the items that frequently occur before and after it in continuous speech, and postulate a representation of these various contexts by multiply crossing and intersecting synfire chains. There are, however, at least five reasons why this strategy is not likely to be successful.

(1) Word sequences span relatively long periods of time, at least ~1 second but usually several seconds. Synfire chains provide precise adjustment of neuronal events that follow each other at short delays, usually in the millisecond range, and occasionally up to 1 second. The synfire model therefore operates at a time scale different from that relevant for word-chain processing. Whereas synfire chains may provide a plausible model of articulatory programs within syllables and words, they cannot provide the mechanism relevant for sequencing words, because this mechanism operates at a *larger time scale*.

(2) The occurrence of a word usually does not allow for good predictions on the particular word(s) that follow(s). Cases in which the occurrence of one particular word predicts, with high probability, the occurrence of a particular complement word, as it is the case for "neither ... nor," represent rare exceptions. Whereas within a word a phoneme is usually followed by 1 of 5 to 20 other phonemes (Harris, 1955), the number of possible successors of a word can be in the order of 10^4. Although a synfire model for phoneme sequences appears feasible, such a model for word sequences would require an astronomic number of chains as a result of the very *large number of possible word sequences*.

(3) The regularities determining word sequences likely operate on more abstract word groups called *lexical categories* (see Chapter 7). Examples of lexical categories are noun (N) or verb (V). The occurrence of a word from one particular category, for example, the noun category predicts with high probability the later occurrence of a member of a complement category – for example, verb (cf. the problem of long-distance dependencies addressed in Section 7.3). However, there is freedom as to which verb to select. It is unclear how a synfire model could realize *lexical categorization*.

(4) When competent speakers are being confronted with new sentences they have encountered before, they may nevertheless judge these new sentences to be correct. This requires generalization from a limited input to new sequences that were not subject to learning. The synfire model leaves open the question how to neuronally implement such *generalization*.

(5) The occurrence of a word of a particular type predicts the occurrence of complement words, but there can be freedom as to at which position the complement follows its antecedent. For example, the occurrence

> of a noun or personal pronoun predicts the later occurrence of a complement verb, but material may intervene between the two, as, for example, in “Peter comes to town,” “Peter, the singer, comes . . . ,” “Peter, the greatest singer in the world, comes . . . ” (see also examples in Chapter 7). A synfire model would not allow for such extremely variable delays.

These points are closely related to the issues discussed in Chapter 7. Although they are generally difficult to account for in a neurobiologically realistic model, the synfire model appears to be unable to solve them. One may speculate whether it might be possible, in principle, to modify the synfire mechanism so that it could operate at a larger time scale and deal with the various problems listed. However, the concept of synfire chains would be stretched in this case, and its most important feature – the temporal precision they provide – would need to be removed [cf. point (5)]. There is no reason for such redefinition of the concept. Rather, it is advisable to characterize the mechanisms necessary for grammar processing and how they differ from the synfire mechanism.

8.4 Functional Webs Composed of Synfire Chains

In Chapter 2, *functional webs* were defined as strongly connected neuron ensembles. This does not have further implications about in which way the assemblies are structured internally. One could envisage these networks to be unstructured lumps of neurons in which average connection probability and strength are high, but in which activity flow is not determined otherwise. In Section 8.2, it was proposed that functional webs may include synfire chains for certain tasks (e.g., for defining articulatory programs). Taking this view, one may still consider cell assemblies to be largely unstructured neuron ensembles in which only subgroups of neurons exhibit well-timed spatiotemporal activity patterns.

An alternative view is the one proposed by Donald Hebb (1949) in his book on cell assemblies. He suggested that the cell assembly is actually a reverberatory chain of neurons, with many reentrant loops through which activity waves can travel repeatedly.

This view about the inner structure of cell assemblies is illustrated in Figure 8.3, which is taken from Hebb (1949). In this diagram, arrows are used to represent neuron groups and the numbers are used to indicate the sequence in which the different neuron groups are activated, given the cell assembly is in a certain state of activity. If the cell assembly is active, this

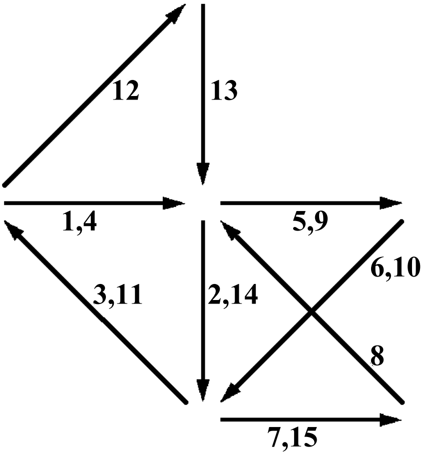

Figure 8.3. Hebb's illustration of the internal structure of a cell assembly. Neuron subgroups of the assembly – represented as arrows in the diagram – were envisaged to become active in a well-timed reverberatory fashion. The numbers indicate the sequence in which the subgroups become active. Neurophysiological observations support the idea of multiply reverberatory circuits (Abeles et al., 1993). Reprinted from Hebb, D. O. (1949). *The organization of behavior. A neuropsychological theory.* New York: John Wiley.

implies that at least some of the neuron groups included in the assembly become active repeatedly. For example, the arrow labeled "1, 4" in the diagram represents the first neuron group to be activated, and it would again be active as the fourth group if the indicated spatiotemporal pattern of activity runs through the cell assembly. According to this view, a cell assembly is conceptualized as a network of neurons with exactly determined spatiotemporal activity patterns. A cell assembly would be a synfire chain with loops, also called a *reverberatory synfire chain* (Abeles et al., 1993). The two terms *cell assembly* and *reverberatory synfire chain* would be synonyms (or, at least, near synonyms; for discussion, see Pulvermüller, 2000).

Because ample redefinitions of a term never facilitates scientific communication and some authors have proposed to distinguish cell assemblies without internal spatiotemporal structure from reverberatory synfire chains (e.g., Braitenberg & Pulvermüller, 1992; Braitenberg & Schüz, 1998), the term *cell assembly* here is restricted to this use. In contrast, the term *functional web* is reserved, in the context of this book, to refer to networks that include reverberatory loops and produce well-timed spatiotemproal patterns of activity (see also definition in Chapter 2). The reverberatory loops are thought to allow the ensembles to retain their activity for some time. In addition, the reverberatory loops in the networks are assumed to be the basis of fast oscillatory activity (Bressler, 1995; Pulvermüller et al., 1997; Tallon-Baudry & Bertrand, 1999). As mentioned in Chapter 4, dynamics of high-frequency cortical responses distinguish words from meaningless pseudowords and may also differentiate between words from different categories

(Pulvermüller, Lutzenberger et al., 1999). One may therefore speculate that these high-frequency dynamics reveal the signature of functional webs, including many reverberatory loops.

8.5 Summary

Synfire chains consisting of sets of neurons connected in sequence are a well-established neurobiological mechanism of serial order. Each synfire chain defines a precise spatiotemporal pattern of neuronal activity. Synfire chains may overlap and cross with other chains while keeping separate the spatiotemporal activity patterns they organize. It is possible that the brain exploits the synfire mechanism for controlling precisely timed movements, including articulations. Serial order of phonemes may be organized by synfire chains built into or attached to functional webs representing words. A mechanism different from that of synfire chains may be necessary for establishing serial order of meaningful language units in sentences.

In summary, a neurobiological approach to serial order in language processing suggests that different mechanisms underlie the processing of phoneme sequences within syllables and words on the one hand and the processing of word and morpheme sequences in sentences on the other hand. Chapter 9 discusses the latter.

CHAPTER NINE

Sequence Detectors

The synfire model discussed in Chapter 8 is one example of a realistic model of serial order in the brain. One may call it *realistic* because it has strong footings in neuroscientific research. Which alternative mechanisms for establishing serial order exist in the nervous system? This chapter reviews a class of serial-order mechanisms different from the synfire chain. It is argued that this type of mechanism may be important for organizing grammar in the brain, and an attempt is undertaken to apply the mechanism to the processing of a simple sentence.

9.1 Movement Detection

As emphasized in Chapter 8, the synfire model realizes a sequence of elementary events "A then B" by direct connections between their neuronal representations, α and β. As an alternative, it is possible to connect a third element to both representations of elementary events. The third element, γ, would become active if sequence AB occurs. The third element would be a *mediator* serving the sequence detection process, which could otherwise be performed by a synfire mechanism as well.

The basic idea for this mechanism of *mediated sequence processing* has been formulated by McCulloch and Pitts (Kleene, 1956; McCulloch & Pitts, 1943). In Chapter 6, the cardinal cells responding specifically to strings of events were called *string detectors*. In modern neuroscientific research, several lines of research have proved similar mechanisms in the nervous system of animals.

Many animals respond specifically to stimuli that move. Therefore, they must be equipped with a mechanism for movement detection. The problem of movement detection shares properties with the serial-order problem, and this becomes obvious in the following formulation. If there are two sensory cells, α and β, looking at adjacent areas A and B of visual space, a moving stimulus first appearing at A and later appearing at B stimulates the neurons α and β sequentially. A third neuron, γ, receiving input from both α and β, may function as a detector of a movement in the AB direction. It should respond to the sequential stimulation of α and β, but not to the reverse sequence. The mechanism yielding sequence sensitivity may involve low-pass filtering of the signal from α, thereby delaying and stretching it over time. Integration of the delayed and stretched signal from α and the actual signal from β yields values that are large when the activation of α precedes that of β, but small values when the activations of α and β occur simultaneously or in reverse order. This mechanism of directional sensitivity was first described in the visual system of insects (Reichardt & Varju, 1959; Varju & Reichardt, 1967). Analogous mechanisms of movement detection by mediated sequence processing were uncovered in the visual cortex of higher mammals (Barlow & Levick, 1965; Hubel, 1995), and a related mechanism of sequence detection exists in the cerebellum (Braitenberg, Heck, & Sultan, 1997).

The cerebellar mechanism appears to be closest to the mechanism of sequence detection suggested by McCulloch and Pitts's model circuits (cf. Chapter 6). Actual delay lines, realized as the parallel fibers of the cerebellum, ensure that precisely ordered sequences of neuronal events can simultaneously stimulate the sequence-detecting cardinal cells, realized as cerebellar Purkinje cells (Braitenberg et al., 1997). This is very close to the mechanism sketched in Figure 6.2. The three input units in the figure would, in reality, correspond to inputs to different parallel fibers. The delay units would not be present in the cerebellar mechanism, but parallel fibers of different length would replace them. The sequence detector cell would be analogous to a Purkinije cell responding only if several action potentials arrive simultaneously in its flat dendritic tree through several parallel fibers. It would appear that many aspects of the mechanisms envisaged by McCulloch and Pitts could be confirmed by modern neurophysiological research (see Heck, 1993, 1995).

Again, the specific feature of the mechanisms of mediated sequence processing is that a sequence of elementary events is detected by a third-party element (labeled γ here) that receives input from the neuronal correlates of the elementary events (labeled α and β). This mechanism of *mediated serial-order processing* is in contrast with the *unmediated* or *direct serial-order mechanisms* as, for example, synfire chains – because it is characterized by the existence of neuronal elements devoted to computing serial-order information and mediating between more elementary units. Please note that the two possibilities – mediated, or indirect, sequence detection based on extra third-party neuronal elements vs. direct sequence detection made possible by the synfire architecture – had already been sketched and contrasted in Section 6.1 (cf. Figs. 6.2 and 6.3, respectively). Meanwhile, it became clear that there is strong evidence for the existence of both mechanisms in the nervous system.

9.2 Sequence Detectors for Word Strings

A basic proposal explored in Chapters (10–12) of this book is that mediated sequence detection may be relevant for processing the serial order of words and morphemes in sentences. However, word-order detection cannot be achieved by exactly one of the mechanisms found in the visual system of arthropodes and vertebrates, or in the cerebellum, because of the time domain differences [cf. caveat (1) in Section 8.3]. As is the case for the synfire chain mechanism, the mechanisms for direction-sensitive movement detection usually apply for delays much smaller than 1 second, whereas substantially longer delays occur between sequentially aligned words and morphemes. Still, however, Barlow and colleagues reported that some neurons in the visual system of vertebrates exhibit rather long decay times (Barlow, Hill, & Levick, 1964), which could be compatible with the detection of sequences spanning tens of seconds. In addition, a model building on sequence detectors fed by word webs can also be subject to all of the points raised previously against a synfire model of word sequencing. Points (2) to (5) in Section 8.3 and also Point (1), are therefore be addressed again. The strategy is to explore what the mediated sequence detection mechanism already well established by neuroscientific research can achieve, and how it would operate at the level of functional webs.

(2) *Number of represented sequences*: One may argue that a model based on indirect sequence detectors for word strings requires a very large number of such detectors, each responding to a particular sentence. However, this is not necessarily so. Like movement detectors, word-sensitive sequence detectors can be assumed to operate on *pairs* of elementary units. If there is a sequence detector for each frequently occurring sequence of two words, the number of necessary sequence detectors can be reduced substantially. Still, the number would be large [but see (3)].

(3) *Categorization*: If a sequence detector γ responds to a sequence "first α_1, then β_1" of neuronal events, it is possible that it responds to a sequence "first α_2, then β_2" as well (where $\alpha_1, \alpha_2, \beta_1$, and β_2 symbolize word webs). By connections to $\alpha_1, \alpha_2, \ldots, \alpha_m$ on the one side and to β_1,

$\beta_2, \ldots, \beta_n$ on the other, γ can be sensitive to activation sequences of elements of groups of word webs – that is, to a sequence of any member of the α group followed by any member of the β group. The α group could, for example, be the lexical category of nouns, and the β group could be the verbs. The sequence detectors could operate on webs representing words and morphemes from given lexical categories.

(4) *Generalization*: Suppose a sequence detector γ is frequently active together with the activation sequence of word webs α_1 and β_1, and develops, by associative learning, strong connections to both so that it finally responds reliably to the sequence "first α_1, then β_1." Additional confrontation with the sequences "first α_1, then β_2" may also strengthen the sequence detector's connections to β_2, and finally, if the activation of α_2 is frequently followed by that of β_1, the α_2 web may furthermore be chained to γ. The "generalization" that the sequence detector is also sensitive to the event "first α_2, then β_2," although this particular sequence may never have been present in the input, follows from this learning history. This type of substitution-based associative learning can account for at least one type of generalization of syntactic rules to novel word strings. It requires that a few standard strings be perceived frequently, and that substitutes replace members of the string selectively so that their representation can be chained to the sequence detector.

(5) *Variable delays*: A sequence detector does not require fixed temporal delays between the activation of the units feeding into it to become active. Motion detectors of the Reichardt–Varju type (Reichardt & Varju, 1959; Varju & Reichardt, 1967) can respond to stimuli moving with variable speed, and in the same way, a functional web fed into by two word webs may respond to their serial activation independently of the exact delay in between their respective activation. A noun–verb sequence detector may therefore become active whenever confronted with one of the strings "Peter comes to town," "Peter, the singer, comes …," or "Peter, the greatest disk jockey of the world, comes … ." Clearly, there must be an upper limit for the delays possible, which, in a Reichardt–Varju inspired model, would depend on the decay times of the word webs and the characteristics of the low-pass filter.

Returning, finally, to Point (1), the time-scale problem, it appears that delays of several seconds do not constitute a principled problem for a model of mediated sequence detection. Again, the decay time and the characteristics of the assumed low-pass filter may determine the time scale on which sequence detection is possible. If functional webs corresponding to words feed into a sequence detector, the activation and decay times of these word webs would likely be crucial. At present, there are no data revealing these dynamics of the neurobiological correlates of words. However, one may allow for educated guesses. Relevant neurophysiological data come from memory cells (see Chapter 2) whose dynamics may reflect the activity status of distributed memory networks they are part of. If this is correct, functional webs could have decay times of 10 seconds or more (Fuster, 1995, 1998b, 2000; Zipser et al., 1993). Assuming that this feature generalizes to the proposed word webs (see Chapter 4), the indirect sequence detectors responding to word-pair sequences may well be assumed to operate on a time scale of several seconds as well.

In summary, the proposal is that mediated sequence processing known from other neuroscientific domains is an important mechanism for syntactic processing. In contrast to the already known mechanisms of sequence detection that operate at the single neuron level, single neurons represent the input and mediate the sequence, the present proposal puts that the same type of mechanism exists at the level of functional webs. Thus, the relevant sequence detectors would be functional webs responding to sequences of neuron populations related to the processing of single words. A sequence detector would become active if the word A_i from a word category A is followed by a word B_j from category B, thereby activating the corresponding functional webs α_i and β_j sequentially. Frequent cooccurrence of words in linear sequences may be an important factor for establishing neuron ensembles specializing in the detection of word sequences. This allows for an economic representation of word-pair sequences, largely independent of the actual delay between the words within a sentence.

In Chapters 10–12, the idea roughly sketched here is more precisely formulated, extended, and illustrated by putative grammar mechanisms. The final section of this chapter may give a first idea how this proposal might connect to syntactic theories and how it differs from them.

9.3 Sequence Detectors and Syntactic Structure

The postulate that word sequences are assessed by sequence detectors leads to a novel view on syntactic processes. The dominating view in linguistics has been that a hierarchical tree of syntactic category representations is built up to parse a sentence and that the individual words of the sentence are attached to the tree as its leaves (see Chapter 7). The tree would have the sentence symbol S as its root, and branches would lead to phrase nodes and

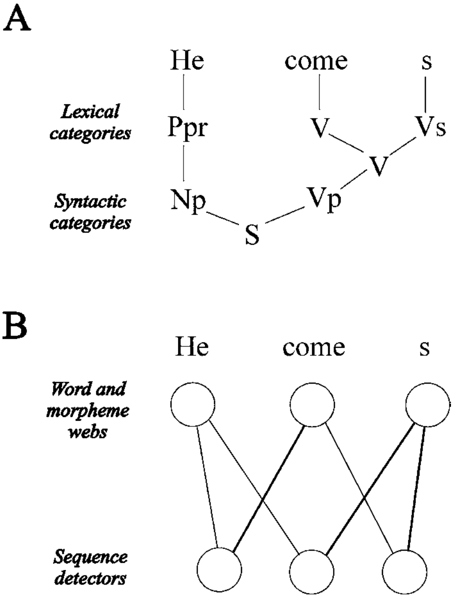

Figure 9.1. (a) A phrase structure representation of the sentence "He comes." Lines represent structural relationships. Abbreviations: Ppn: personal pronoun; V: verb; Vs: verb suffix; Np: noun phrase; Vp: verb phrase; S: sentence. (b) Putative neuronal circuit based on mediated sequence detection. Circles represent functional webs. Labels close to circle indicate the morphemes represented by word webs (upper line) and the sequences of lexical category members sequence detectors are assumed to be sensitive to (lower line). Thin and thick lines represent qualitatively different neuronal connections between sequence detectors and word/morpheme webs, respectively.

to lexical category nodes. Another example syntactic tree for a maximally simple sentence, or sentence fragment, is presented in the upper diagram of Figure 9.1. The phrase structure of the string "He comes" is used as an example. As illustrated in Chapter 7 with different examples, this sentence can also be broken down into noun phrase (Np) and verb phrase (Vp). The noun phrase is, in this case, realized as a personal pronoun (Ppr), and the verb phrase as verb stem (V) plus verb suffix (Vs).

The tree representation per se has the disadvantage that it cannot capture one type of long-distance dependency called *agreement* (see Chapter 7). The relationship between the sentence-initial pronoun and the sentence-final suffix requires an extension of the concept of a two-dimensional tree structure. As detailed in Chapter 7, linguists have proposed supplementary mechanisms to model the interdependence of these elements, which, in the present example, agree in person and number. The most popular approach proposes that features of the items be transported through the branches of the tree to mediate between its leaves. A disadvantage of this strategy is that it is not economical: Two mechanisms – tree construction and within-tree transport of features – are postulated, although one may speculate that one mechanism could equally well represent and process the multiple relationships between the constituents of the sentence (see Chapter 7).

A syntactic model based on sequence detectors replaces the tree construct by a set of neuronal elements mediating between word webs (Figure 9.1, bottom diagram). Separate sequence detectors responding to word pairs, in example (a) to the pronoun–verb sequence, (b) to the verb–verb suffix sequence, and, in the very same way, (c) to the pronoun–verb suffix sequence, are envisaged to be activated by the word string. The activation of these three sequence detectors would also represent and process structural information about the word string. Each sequence detector would represent and, given it is active, store the information that a particular pair of syntactically related morphemes has recently been present in the input.

At first glance, this approach may appear more economical than the syntactic-tree-and-feature-transport approach because it postulates only one unified mechanism, mediated sequence detection, which, in this case, can replace two different mechanisms, for subordination and agreement in syntactic trees.

Psychological phenomenon may provide additional support for the neurobiologically inspired idea that mediated sequence detection operating on categories of words could underlie syntactic processing. An important observation is that previously perceived syntactic structures are being imitated in subsequent verbal actions. This phenomenon attributed to a mechanism dubbed *syntactic priming* occurs with above-chance probability in both conversations and controled experiments (Bock, 1986; Bock, Loebell, & Morey, 1992; Pickering & Branigan, 1999). The phenomenon is independent of whether or not the two sentences share words. A double object sentence as a prime ("showed the children the pictures") yields later production of other double object sentences ("gave the butcher the knife"), and a similar priming effect can be observed for the prepositional object paraphrase ("showed the pictures to the children"). It has been pointed out that certain aspects of syntactic priming are difficult to explain on the basis of some of the available syntactic theories (for discussion, see Pickering & Branigan, 1999).

It is obvious that imitation of sequences of words from similar lexical categories can be explained easily and straightforwardly based on the notion of sequence detectors operating on categories of word representations. If sequence detectors play a role in both the detection and production of strings of words from given lexical categories, the explanation of syntactic priming can be based on primed sequence detectors. Priming of these neuronal webs by an incoming sentence enhances the activity level of these neuronal units, thus later enhancing the probability that similar word sequences will be produced.

There cannot be any doubt that networks made up of neurons can realize important aspects of the serial order of events. It is, nevertheless, important to point to some of the neurocomputational research that investigated in detail the mechanisms discussed here. There is a long history of work exploring the capabilities of associative networks, which has been sparked by theoretical proposals (e. g., McCulloch & Pitts, 1943) and empirical results (e.g., Reichardt & Varju, 1959). Willwacher (1976, 1982), for example, presented an early implementation of a single-layer network capable of learning and retrieving letter sequences, and Buonomano (2000) recently showed that a variety of precise delays between events can be learned and represented in an associative network consisting of excitatory and inhibitory neurons that are organized in one neuronal layer. Some researchers have included into their simulations much detail about the specific features of the sequences under study, such as complex grasping or walking movements (Cruse & Bruwer, 1987; Cruse et al., 1995), and about neuronal responses as revealed by neurophysiological investigation (Kleinfeld & Sompolinsky, 1988).

Apart from single-layer associative networks, more complex networks have been used with some success. Elman used an architecture that includes hierarchically organized layers, one of which is reciprocally connected to an additional so-called memory layer where information about past events can accumulate. This architecture proved particularly fruitful for modeling serial order of language elements (Elman, 1900; see Chapter 6). Dehaene used a three-layer model, including one layer where sequence detectors could develop very similar to the ones discussed here in the context of syntactic processes (Dehaene, Changeux, & Nadal, 1987).

Despite these successes in modeling serial order relationships in neural models, it should be kept in mind that the successful application of a network to a problem of serial order does not clearly imply that the relevant mechanisms on which sequence production or detection is based have been uncovered. For some simulation approaches it remains to be shown whether the crucial mechanism is direct sequence detection by delay lines or, as an alternative, mediated sequence detection relying on separate neuronal units devoted to the processing of serial order information. This can be decided by looking closely at the behavior of individual neurons included in the network.

The model of sequence detection discussed here makes specific predictions on the outcome of neurophysiological experiments that have, as to the author's knowledge, not been carried out yet. The considerations on syntax offered in this section would suggest that it might be advantageous to have neuronal units available that respond specifically to a sequence of events A and B, but that their response is largely independent of the delay. A further prediction might be that the relevant delays range between 0.2 seconds and tens of seconds. The model discussed here would suggest that such sequence detectors responding to specific word sequences would be particularly common in the left perisylvian cortex. Furthermore, the neurobiological approach may provide a brain-based explanation of neurophysiological and metabolic changes in brain activity related to the processing of syntactic information (Caplan, 2001; Friederici, Pfeifer, & Hahne, 1993; Hagoort, Brown, & Groothusen, 1993; Indefrey et al., 2001; Moro et al., 2001; Neville et al., 1991; Osterhout & Holcomb, 1992). This issue will be raised again in Chapter 13.

After the illustration of how sequence detectors may provide a basis for grammar, it now becomes important to formulate the proposal in a more general way.

CHAPTER TEN

Neuronal Grammar

Large, strongly connected groups of neurons were proposed to form the neurobiological substrate of higher cognitive processes in general and language in particular. If the reader wishes, the ultimate answer to the question of language, the brain, and everything was suggested to be neuronal ensemble. Different authors define terms such as neuron ensemble, cell assembly, and neuronal group in different ways, and therefore a new term, functional web, was proposed and its meaning clarified (see Chapters 2, 5, and 8). There is support for the concept of functional webs from neurophysiological and neuroimaging experiments on language and memory (see Chapters 2 and 4). In this chapter, the notion of a functional web is used as a starting point for a serial-order model. The elements of this model are called *neuronal sets*. Neuronal sets are functional webs with additional special properties that are relevant for serial-order processing. Neuronal sets can represent sequences of words and are then called *sequence sets* (or alternatively, *sequencing units*, or *sequence detectors*). New terms are introduced to distinguish the entity that has empirical support (functional web) from the theoretical concept developed (neuronal set).

In this chapter, the notions *neuronal set* and *sequence set* are explained and applied to introduce a putative basic mechanism of grammar in the brain. Properties of functional webs are first briefly summarized, and then the concept of a *neuronal set* is defined as a special type of functional web. The process of the detection of word sequences is spelled out in terms of the ordered activation and deactivation of neuronal sets. The possible neurobiological correlate of lexical categories is discussed.

10.1 The Story So Far

A functional web is a widely distributed group of cortical neurons that are strongly connected to each other (see Chapter 2). Because of its strong internal connections, the web is assumed to act as a functional unit, that is, stimulation of a substantial fraction of its neurons lead to spreading activity and finally to full activation of the set. This full activation is called *ignition*. Although after an ignition has taken place refractory periods of neurons and other fatigue effects again reduce the excitation level of the set, activity in the web is proposed not to die out completely. The strong internal connections still allow activity to be retained in the web. A wave of activity may reverberate in the set, as proposed by Hebb (see Chapter 8). Results of studies of cognitive processes using noninvasive physiological recordings are consistent with the view that ignition and reverberation are successive processes reflected in stimulus-evoked cortical potentials and induced high-frequency activity (see Chapter 4).

It is important to note the differences between the concepts of reverberation and ignition. Ignition is a brief event, whereas reverberation is a continuous process lasting for several seconds or longer. Ignition of a functional web involves all of its neurons, or at least a substantial proportion of them, whereas reverberation can be maintained by small neuron subgroups within the set being active at given points in time. Ignition does not imply a fixed spatiotemporal order of neuronal activity. In contrast, reverberation is characterized by a fixed sequence of neuron activations, a defined spatiotemporal pattern of activity within functional webs. Ignition would be the result of the overall strong connections within a web, and reverberation would be made possible by its strongest internal links envisaged to provide a preferred highway for spreading neuronal activity, so to speak. Importantly, two distinct types of activation of a web are postulated. In addition, there is the possibility that the web is in an inactive state – that is, that it stays at its resting level. Therefore, a functional web can have three activity states: It can rest, reverberate, or ignite.

10.2 Neuronal Sets

Neuronal sets are functional webs characterized by a greater variety of activity states.

If different functional webs are connected to each other, the excitatory processes of ignition and reverberation in a given ensemble likely influence other webs directly connected to it. Through connections between webs, which are assumed to be much weaker, in the average, than their internal connections, one set of neurons can have an activating effect on the other. This between-web stimulation does not necessarily mean that there is a long lasting effect in the second set. A continuously reverberating and autonomous wave of activity is not necessarily generated in the stimulated set, and an ignition only takes place if the activation from outside is substantial, above an ignition threshold. Subthreshold activation of a web as a result of input from one or more other sets is called *priming at the neuronal level*. Neuronal sets are conceptualized as functional webs that can exhibit priming. Thus, neuronal sets are characterized by four different types of possible activity states: ignition (I), reverberation (R), priming (P), and inactivity (0).

To clarify the main differences between terms used in this book to refer to neuron groups of different kinds, Table 10.1 gives an overview. The terms *neuron group, cell assembly, functional web, neuronal set*, and, finally, *sequence set* or *web* refer to subcategories with increasingly narrower definitions.

The term *priming* must not be mixed up with priming at the behavioral level (e.g., semantic priming; see Chapter 5), although it is possible that the neuronal mechanism is the cause of the behavioral phenomena (Milner, 1957). Because later chapters deal with neuronal mechanisms, the term *priming* usually refers to the neuronal level unless otherwise indicated.

For a network formed by many neuronal sets, with stronger and weaker connections between the sets, it appears plausible to assume that different activity states of one set may differentially affect other connected sets. Because ignition is a substantial excitatory process, it may affect other directly connected sets, or *neighbor sets*, regardless of how strong connections are between the sets. In contrast, the less powerful process of reverberation, to which less neurons contribute at any point in time, may be visible in neighbor sets only if between-set connections are strong. Ignition is assumed to prime connected neighbor sets, whereas reverberation is assumed to prime neighbor sets only if the connection between the sets is strong.

In the state of reverberation, a neuronal set is similar to a McCulloch–Pitts neuron of threshold one with one self-connecting loop. However, as already mentioned in Chapter 2 and later, neurophysiological evidence does not support the notion of a neuronal element exhibiting constantly enhanced activity over longer periods of time (Fuster, 1995). Instead, activity levels of memory cells in many cases exhibit an exponential decrease of activity. If the memory cells reflect the activity state of larger sets, it makes sense to assume that also the larger sets show exponential decline of activity with time if the set is in its reverberatory state. The activity level A of a neuronal

Table 10.1. Overview of the different terms used to refer to different types of neuron groups. Terms, the most crucial features of their referents, and aspects of their underlying putative mechanisms are summarized. (For explanation, see text.)

Concepts of Cell Assemblies

Term	Meaning	Mechanism
Neuron group	Selection of neurons (may not be connected)	
Neuronal ensemble	Selection of neurons that	
Neural assembly	Are strongly connected to each other	Associative learning in autoassociative memory
Cell assembly	Act as a functional unit, ignite	Strong connections between neurons
Functional web	Selection of neurons that	
	Are strongly connected to each other	Associative learning in autoassociative memory
	Act as a functional unit, ignite	Strong connections between neurons
	Maintain activity, reverberate	Reverberating synfire chains connecting neurons
Neuronal set	Selection of neurons that	
	Are strongly connected to each other	Associative learning in autoassociative memory
	Act as a functional unit, ignite	Strong connections between neurons
	Maintain activity, reverberate	Reverberating synfire chains connecting neurons
	Can be primed	Efferent connections to the neurons
Sequence set	Selection of neurons that	
Sequence detector	Are strongly connected to each other	Associative learning in autoassociative memory
Sequence web	Act as a functional unit, ignite	Strong connections between neurons
	Maintain activity, reverberate	Reverberating synfire chains connecting neurons
	Can be primed	Efferent connections to the neurons
	Respond specifically to sequences	Indirect sequence processing

set S_i at time t may be described by Eq. (1).

$$\begin{aligned} A(S_i, t) &= A(S_i, t_0)\, e^{-c\Delta t} \\ &= A(S_i, t_0)\, e^{-c(t-t_0)} \end{aligned} \tag{1}$$

where t_0 is the point in time when the set has received excitatory input from neurons outside the set, Δt is the time difference between t and t_0, $A(S_i, t_0)$ is the amount of activity it exhibited at t_0, and c is a constant determining the slope of the exponential decrease of activity (c can be assumed to be 1 here). The larger Δt, the smaller the exponential term $e^{-c\Delta t}$. In essence, when reverberating, the neuronal set continuously loses activity, and this activity decline is assumed to be exponential.

As in many network simulations (see Chapter 6), the time continuum is sliced into discrete time steps. This is done for illustration and simulation purposes. In reality, the exponential decrease of activity is probably continuous rather than in discrete steps.

If a set receives excitatory input from outside, an ignition may result. If the input is not sufficient for reaching the ignition threshold, the set is merely being primed. In contrast to ignition and reverberation, priming of a given set depends on input from other sets. Priming therefore terminates as soon as the input has ceased.

One may question this assumption, because excitatory postsynaptic potentials are known to last for tens of milliseconds after an action potential has arrived (see Chapter 2). However, activity dynamics of neuronal sets involved in processing serial order of words would probably operate at a much longer time scale (see Chapter 8). If the time steps considered in a model of sequence detection are on the order of 30 or so milliseconds, excitatory postsynaptic potentials can probably be considered as momentary phenomena.

Again, the momentary process of priming is distinguished from the long-lasting process of reverberation. Certain forms of behavioral priming (e.g., repetition priming) would therefore need to be attributed to reverberation rather than to priming at the neuronal level. From a neurobiological point of view, the concept of priming as it is used in psychology may correspond to a variety of different neuronal mechanisms (see Mohr & Pulvermüller, in press, for further elaboration).

Under priming, the various excitatory inputs to a given set from its active neighbors summates at each time step. Inputs from neighbor sets that reverberate or ignite are being received at the same time through different input fibers of the neurons of the set, and this can lead to spatial summation in its neurons. One may also say that the neuronal set as a whole summates the various simultaneous inputs so that its overall activity level increases. If additivity of activity is assumed, the following equation gives the activity level of a set S_i at time t, which received input from j other sets S_j one time step before:

$$A(S_i, t) = \Sigma_j A(S_j, t - 1) \tag{2}$$

where $A(S_j, t - 1)$ indicates the amount of excitatory stimulation a set S_j sends to set S_i at a given time step $t - 1$ and which arrives at S_i at t. Although Eq. (2) may give a rough idea about the real processes, these are likely more complex because the underlying assumption of additivity is idealized.

If an already reverberating set receives stimulation from outside, the excitatory effects of the reverberation itself and the inputs from other sets are again assumed to summate.

$$A(S_i, t) = A(S_i, t_0)^{-c(t-t_0)} + \Sigma_j A(S_j, t - 1) \tag{3}$$

This formula is one way to express that two main sources of excitation of a neuronal set are (a) its internally reverberating excitation and (b) the input to the set through afferent connections from other neurons – that is, through priming. As an alternative to mere summation, the activity provided by reverberation and priming may interact in a more complex manner.

As in the artificial neuron-like machines proposed by McCulloch and Pitts, the integrated activity can be used as a criterion for the decision about fully activating a neuronal set. If the activity level summed over space and time exceeds a certain value, the ignition threshold θ, then the set ignites. After its ignition (I), and everything else being equal, the set starts to reverberate again at a given reverberation level, R_1. To state this briefly:

$$A(S_i, t) > \theta \Rightarrow A(S_i, t + 1) = I \Rightarrow A(S_i, t + 2) = R_1 \tag{4}$$

The double arrows denote causality and temporal succession. One time step after the threshold is reached, an ignition occurs which is, again after one time step, followed by the reverberation process.

In essence, neuronal sets are assumed to have qualitatively different possible activity states. At each point in time, a set is characterized by one activity state from the set AS of possible activity state types:

$$AS = \{0, P, R, I\} \tag{5}$$

Again, 0 is the resting level; I is full activation or ignition; R represents reverberation, which follows ignition; and P is priming caused by the ignition or reverberation of a different but connected neuronal set.

Whereas 0 and I refer to single activity states, P and R denote sets of possible states. As Eqs. (1) and (2) imply, there are different levels of reverberation and priming. The set R of reverberatory states is defined by the exponential function Eq. (1) and, therefore, there is, in principle, an unlimited number of possible activity states.

$$R = \{R_1, \ldots, R_i, \ldots, R_n\};\, R_1 < \theta,\, R_n > 0 \tag{6}$$

There is no upper limit for the number n of the possible reverberatory states R_i because decline of activity is continuous, and R_i is therefore from the set of positive real numbers. Thus, theoretically speaking, there are uncountably many possible states R_i. From this, it follows that a neuronal set cannot be considered a finite state device because a finite state device is defined as having only a finite number of possible states.

Nevertheless, a set of neurons with infinitely many possible activity states is unlikely to exist in reality. The noise available in all biological systems cancels the minor differences between similar states so that, in fact, it is only possible to distinguish a finite set of states. The influence of the noise depends on its ratio to the signal. If a neuronal set consisted of a very large number of neurons, each stimulated by statistically independent noise, it would be possible to minimize or even cancel the effect of random noise to each individual neuron by averaging activity over all elements of the set at each point in time (see Chapter 2). Given good signal-to-noise-ratios are being computed, the number of possible activity states can therefore be assumed to be high, although not infinite. The number of distinguishable states of reverberation of a set is limited by the signal-to-noise ratio of the system.

A similar point as the one made above for reverberation can be made for levels of priming as well. As implied by Eq. (2), there is no principled upper limit for the number of members of the set P of states of priming.

$$P = \{P_1, \ldots, P_i, \ldots, P_n\};\, P_1 < \theta,\, R_n > 0 \tag{6}$$

Although the number of inputs to a neuronal set is limited, n can be high because each input can be strong or weak, depending on connection strength and the activity level in the stimulating set. Because this activity level of the stimulating sets can vary, the input to the primed set varies as well.

Although there are theoretical and empirical reasons to conceptualize neuronal sets as networks with properties as described by Eq. (1) to Eq. (5), certain simplifications appear necessary to illustrate the functioning of grammars composed of neuronal sets. In this and the following two chapters and the related Excursuses, grammar circuits are introduced. These are kept as simple as possible. The circuits and their functions are designed to provide comprehensive examples of the computational perspectives of grammars formulated in terms of neurons. The following simplifications are made to allow for straightforward and well-intelligible illustration:

(1) Only a few states of reverberation (labeled R_1, R_2, R_3, etc., or, alternatively, R, R', R'', etc.) and priming (P_1, $P_2, \ldots$, or P, $P', \ldots$) are mentioned explicitly in the examples that follow.

(2) Continuous activity decrease with time [Eq. (1)] is ignored except if a new level of activity is being introduced. R_1 denotes the highest activity level in a network at a given point in time (irrespective of its absolute value); R_2 denotes the second highest activity level, just one step below; R_3 refers to the third-highest levels; and so on. The same applies for levels of priming.

(3) Spatiotemporal summation of activity and the criterion for ignition is simplified.

Again, this is done to maximize intelligibility of the formalisms and make computations less complex and therefore easier to illustrate. The exponential activity decline of neuronal sets with time is proposed to be crucial for storage of serial-order information.

10.3 Threshold Control

In Chapter 5, it was argued that for a cortex-like network to operate properly, it is necessary to assume a regulation mechanism that keeps the activity level within certain bounds. A feedback regulation mechanism can be assumed to adjust the thresholds of individual neurons by way of providing or removing background activity to move the cortical activity level close to a target value. This mechanism has been called the *threshold control mechanism* (Braitenberg, 1978a; Braitenberg & Schüz, 1998; cf. Chapter 5).

Here, it is assumed that the regulation mechanism detects fast activity increases and removes background activity to the cortex if a substantial activity increase is detected. If the level of activity is generally low, the mechanism provides additional excitation. As emphasized in Chapter 5, other criteria for activation of the regulation mechanism can easily be imagined, and at this point, it is not clear which of the several possibilities is actually realized in the real brain. However, because it is evident that a feedback regulation mechanism of some kind is necessary, one of the several options must be chosen for realistic modeling.

A criterion based on activity increase can be realized in a network without difficulty. The global activity state summed over all neurons in the network at a time point t could be compared to global activity at an earlier time (e.g., one time step earlier, at $t-1$). Only if the difference exceeds a certain value is inhibitory input to the network provided. An even simpler criterion for activity increase can be introduced on the basis of the following consideration. The most extreme activity increase is probably present in a network of neuronal sets if one or more sets ignite after being in the resting state 0. Therefore, such fast activity increase may be assumed to cause the regulation mechanism to become active and initiate a process of global inhibition affecting the entire network.

Compared to the ignition of an inactive set, the activity change is smaller if an already primed or reverberating neuronal set ignites. Ignition implies simultaneous activity of the majority of the neurons in the set. The number of neurons in a preactivated set that become active during an ignition (in addition to those already being active) is therefore smaller compared to a previously inactive set that ignites. It is assumed that full ignition of a neuronal set at rest activates the threshold control mechanism, whereas the smaller activity increase induced by the primed ignition of an already preactivated or primed set does not. The symbol $I^\wedge$ is used to refer to a full ignition that subsequently activates threshold regulation, whereas I denotes ignition without any consequences for the threshold regulation machinery.

The expression *preactivity of a set* is used to refer to any activity state different from rest (0) and ignition (I); that is, either reverberation (R) or priming (P).

A full ignition $I^\wedge$ leads to global inhibition in the network. This is realized as follows: When $I^\wedge$ occurs in one set, all preactivity states of other sets – that is, reverberation and priming levels – is diminished. R_1 and P_1 are changed to R_2 and P_2, and, generally, all states n are changed to $n + 1$. Recall that the number gives the rank in the hierarchy of activity states, 1 being higher than 2. A full ignition $I^\wedge$ reducing preactivity states therefore changes all states R_n and P_n to R_{n+1} and P_{n+1}, respectively.

If inhibition happens regularly, each reverberating set loses activity over time. Thus, a threshold regulation mechanism becoming effective with constant frequency may cause constant activity decrease, as described by Eq. (1), in preactivated neuronal sets. Furthermore, a global inhibitory mechanism could guarantee that the slopes of activity decline are the same in different sets, an assumption relevant for the pushdown memory mechanism focused on in Chapter 12 and Excursus E5 (see also Pulvermüller, 1993).

As mentioned, a threshold regulation mechanism can also have an excitatory effect. It is assumed that if activity levels in all reverberating and primed neuronal sets are generally low, additional global excitation is provided by the regulation mechanism. All preactivity levels therefore become higher.

Because the formulation of preactivity states refers to relative activity levels (e.g., "R_1" meaning "best activated set," not a particular absolute activity level), the activating function of the regulation mechanism is built into the formulation, so to speak. One or more sets is always at the highest level of reverberation and priming. A global increase of activity shows up only as relabeling of activity levels if there is no neuronal set in the network exhibiting the highest levels of reverberation and priming (R_1 and P_1). In this case, the numbering is adjusted so that the highest activity level is 1.

10.4 Sequence Detection in Networks of Neuronal Sets

Based on which wiring would a neuronal unit respond to a sequence AB of events but not to BA? As outlined in Chapter 9, one could have strong connections from an input unit α, specifically responding to the occurrence of an A in the input, to a sequence detector, and between the sequence detector and β, the neuronal unit responding to the occurrence of a B in the input. Through these strong bottom-up connections, information about activity of both α and β could influence activity of the neuronal set γ mediating sequence detection.

A few terminological issues should be mentioned at this point. Neuronal sets that respond specifically to the ordered sequence of activations of two other sets are called *sequence sets*. The words *sequence detector* and *sequencing unit* are also used to refer to neuronal sets mediating sequence detection. If the sequence set γ only responds to the event "first α, then β," but not to the reverse sequence, the event A (to which α responds) is called the *earlier event* and the event B (to which β responds) is called the *later event*.

The issue of mediated sequence detection has already been addressed in the context of neural networks (Chapter 6) and brain mechanisms of serial order (see Chapter 9). As pointed out, there are different mechanisms available for realizing sequence detection. In McCulloch–Pitts networks, sequence sensitivity is established by connections, from event detectors α and β to the sequence detector γ, with different conduction delays. Delay neurons were introduced to guarantee the delay (see Section 6.1). The Reichardt–Varju theory proposed low–pass filtering of the signal from one of the event detectors (responding to the earlier event) as a possible mechanism crucial for sequence detection (Egelhaaf, Borst, & Reichardt, 1989; Reichardt & Varju, 1959). In this case, the delay of the earlier signal would be provided by a mathematical transformation carried out by nerve cells.

A further possibility is that sequence sensitivity is merely the result of connection strength between event and sequence detectors, and to the activity dynamics of the former. Assume, for example, that the event detector β responding to the later event B has stronger connections to the sequence detector γ than α which responds to the earlier event A. Because of the stronger $\beta\gamma$ connection, any activity in β excites γ more strongly than activity in α. Assume further that α and β are characterized by the same exponential decline function of reverberating activity [Eq. (1)]. Now, if A occurs before B, the declined small signal through the weak $\alpha\gamma$ connection summates with the full strong signal through the $\beta\gamma$ connection. This sum is greater than the $\beta\gamma$ input alone. If B occurs before A, the declined strong signal through the $\beta\gamma$ connection summates with the not degraded weak signal through the $\alpha\gamma$ connection. The latter sum is smaller than the $\beta\gamma$ signal as soon as the degradation of the strong input through $\beta\gamma$ is larger than the full weak input through $\alpha\gamma$. Thus, if the decline function is steep compared to the difference between the strong and the weak full signals, the summed input to γ is greater if A occurs before B than in response to the reverse sequence of events. Direct connections from both α and β to γ are sufficient for providing γ with serial-order information. The threshold for activating γ can be adjusted so that it responds to the activation sequence "first α, then β," but not to the reverse sequence.

In summary, different mechanisms may mediate the sensitivity to serial-order information of neurons and sets of neurons. Because the issue of which of the alternative mechanisms underlie grammar processing in the human brain is still far from being addressed directly by empirical research, it suffices to postulate a mechanism of mediated serial-order detection carried out at the level of neuronal sets (see Chapter 9). The exact details of how this mechanism is realized are left for future research.

Information flow from event to sequence detectors is required for mediated serial-order detection. This could be implemented by unidirectional connections from word webs to sequence sets. However, bidirectional connections between neuronal sets is always assumed to be the default in the present context. The motivation for this is two-fold: One reason comes from neuroanatomical and neurophysiological studies (see Chapters 2 and 8). If long neuronal connections in the cortex are considered the substrate of grammar-relevant mechanisms, one-directional connections are probably the exception rather than the rule. If two cortical areas are connected to each other in one direction, they usually exhibit the reciprocal connection as well (Pandya & Yeterian, 1985; Young et al., 1995). At the level of local cortical circuits, Abeles and colleagues (1993) found evidence for reverberatory loops, a fact implying bidirectional rather than unidirectional links between many of the neurons in the circuit. Because neuroanatomical studies indicate that most connections, local and between areas, are reciprocal, unidirectional connections between large distributed neuronal sets are less likely (see the reciprocicity principle in Chapter 2). The default assumption should be reciprocal connections. Still, the reciprocal connections may be asymmetrical; that is, stronger in one direction than in the other. This is taken into account later in this chapter.

The second reason for assuming bidirectional links is psycholinguistic in nature: It appears that when listening to a sentence, the first words usually lead one to generate hypotheses about the next. In some cases, however, words occurring later in a sentence help disambiguate or fully understand earlier sentence parts. The verb in "Betty *switch*ed" can be understood in different ways, and it is only after perception of later sentence parts that the appropriate reading can be determined. The final part of the sentence, "her shopping mall," would lead to one interpretation of the verb, whereas if the final part of the sentence is "her TV off" would elicit another. This can be modeled by backward connections from sequence sets to the word webs.

For these reasons, neuroscientific and psychological in nature, one may prefer having activity spread in both directions, back and forth, through the representation of a sentence being heard, parsed, and understood. Reciprocal connections are therefore assumed to connect word webs to sequence sets and sequence sets among each other.

The reciprocal connections between word webs and sequence sets is assumed to be asymmetrical. The proposed connection in one direction would be stronger than the connection in the other. If two sets are connected reciprocally but one of the sets usually becomes active before the other, it may be that the connections from the first to the second set become stronger than the links back from the second set to the first. This is an unavoidable conclusion if the following assumptions are being made: (a) Hebbian correlation learning takes place, that is, synaptic strengthening is strongest if simultaneous pre- and postsynaptic activity occurs; (b) all conduction times are non-zero; (c) some conduction delays equal frequent delays between ignitions of two sets (Miller, 1996). Together, these assumptions motivate the assumption of asymmetry in reciprocal connections between neuronal sets.

In this view, information about possible input sequences is represented in a network (i) by links between neuronal sets and (ii) by the direction in which connections are stronger, assuming that the connection is strong in one direction and weak in the other.

Distinguishing only between "strong" and "weak" connections is certainly another simplification. Strength of synaptic connections in the brain can take a large number of different values. However, this minimal distinction between connection strengths allows for complex syntactic processing in neuronal grammar, as exemplified later. The additional advantages of assuming more degrees of freedom for connection strengths is mentioned occasionally in this text, but is not explored systematically.

Furthermore, the assumption that all reciprocal connections between word webs and sequence sets are asymmetric may turn out to be too strong.

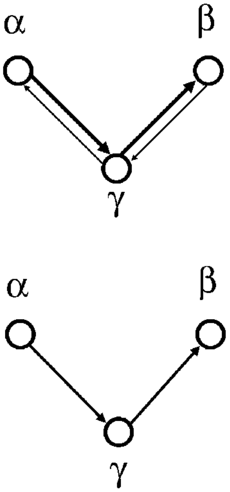

Figure 10.1. Two input units or word webs connected to a sequence set. All circles represent neuronal sets that can be primed, can ignite and reverberate after their ignition. (Top) Strong connections link input unit α to the sequence detector, and the sequence detector in turn to set β. Weak connections run in the opposite direction. The sequence detector becomes active if it is stimulated twice at a time.

In addition, there may be neuronal sets that are sensitive to both possible sequences of two events, AB and BA. These neuronal sets are realized as reciprocal symmetric connections to their input units. This kind of And detector may make it easier to model languages in which there is no preferred order of constituents.

The upper diagram in Figure 10.1 presents a sequence set connected with two input units or word webs. This network is assumed to be capable of sequence detection, that is, it exhibits a certain activity pattern that can be taken as an indicator that a string in the input is consistent with the network structure. In linguistic terms, this may be similar to the judgment that the string in the input is grammatical or well formed. This network differs from those discussed in earlier chapters because all circles in Figure 10.1 denote neuronal sets that can reverberate, ignite, and be primed by other sets. In earlier diagrams, circles represented neurons, local clusers of neurons, or cell assemblies with less sophisticated activity dynamics (see Table 10.1).

In the upper diagram of Figure 10.1, thick and thin arrows are used to indicate relatively strong or weak connections, respectively. The weaker connections make the diagrams more complex, in particular if numerous sets and connections are part of it. The weak connections are therefore omitted, as in the lower diagram of Figure 10.1, which is supposed to have the same meaning as the upper diagram. The same difference exists between the upper and lower diagrams in Figure 10.2. In the diagrams at the bottom, and in all subsequent figures, arrows indicate that there is a strong connection in the direction of the arrow and a weaker connection in the reverse direction. Double arrows may indicate symmetrical connections.

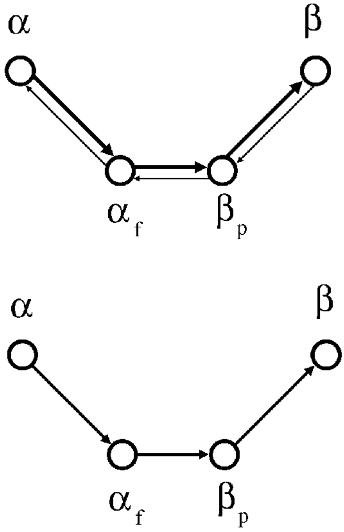

Figure 10.2. Input units connected through two sequencing units: a *forward-looking* unit α_f and a *backward-looking* unit β_p. Sequencing units are duplicated to make formulation of neuronal grammars easier.

In summary, sequence sets may specifically respond to serially ordered ignitions of two neuronal sets corresponding to words (word webs). The mechanism of mediated sequence detection requires information flow, and therefore connections, from input units (e.g., word webs) to sequence detectors (e.g., sequence sets). The sequence detecting set and the event detectors are assumed to be linked reciprocally to each other; however, with asymmetrical weights. Stronger connections are assumed in the direction of the usual activity flow, from the event detector of the earlier event to that of the later event, compared to the respective links in the opposite direction. Although it is clear that the mechanism sketched here is in need of both further theoretical elaboration and empirical support, it nevertheless becomes possible to explore its function in an envisaged grammar network.

10.5 Activity Dynamics of Sequence Detection

The basic idea put forward here is that word webs and the sequence sets connected to them can respond in a specific manner to a sequence of words and fail to do so in response to the same words presented in a different order. The mechanism this is grounded in is the response of the sequence set to sequences it is prepared to detect. If the sequence set γ is connected to both the early event's web α and to the late event's web β, and if γ needs two inputs to become active, ignition of α may first prime γ, so that the ignition of β can later fully activate γ. The primed ignition of the sequence set would then be the neuronal indicator of the fact that the input was in accord with the network structure.

Before examining how a network composed of word webs and sequence sets would operate, it is necessary to introduce a few assumptions about activity dynamics within neuronal sets and activity exchange caused by connections between neuronal sets.

Each word or morpheme in the input activates its input unit or word web that, as a consequence, ignites and thereafter stays in the state of reverberation. As already mentioned, ignition and reverberation of one neuronal set are assumed to have different priming effects on other sets connected directly to the igniting or reverberating circuit. Reverberation causes priming only through strong connections. Ignition causes activity changes through all connections, strong and weak.

Ignitions are assumed to spread from input units to sequence sets, but only to preactivated ones. Thus, to activate a sequence set, it first needs to be primed, or be in the reverberatory state, and then a second strong activating input caused by an ignition must reach it. The latter assumption is somewhat similar to the assumption of a threshold of two of McCulloch–Pitts neurons. Summation of activity already present in a web and the afferent input to it has been specified by Eq. (3). Clearly, if preactivity is low, the second input to the sequence set may not cause it to reach the ignition threshold. Therefore, ignitions are allowed to spread only to a sequence set if it reverberates or is being primed at high levels of R or P. These assumptions are elaborated further later.

The reader is invited to consider how the network in Figure 10.1 operates. To this end, the changes of activity states in the course of a sequence of inputs are illustrated.

Suppose the sequence AB is in accord with the network structure, whereas BA is not. An input sequence in accord with a network structure is called a *congruent input*, whereas a sequence without corresponding wiring in the network, and not otherwise represented, is called an *incongruent input*. This is analogous to the linguistic classification of strings as grammatical, or well formed, vs. ungrammatical, or ill formed. In contrast to the linguistic distinction, which relies on judgments of native speakers of a language, the terms *congruent* and *incongruent* are defined on the basis of putative neuronal circuits. In the circuit in Figure 10.1, the following processes may occur and contribute to sequence detection.

Network response to a congruent input AB is as follows:

(1) The word A occurs in the input, and the word web (or input unit) α ignites.

(2) Because α was not primed and thus exhibits unprimed or full ignition $I^\wedge$, the threshold regulation mechanism is activated. However, at this point, there is no other activated set, and therefore threshold regulation has no effect. All other sets stay inactive.
(3) The ignition of α temporarily primes its connected sequence detector and input unit. This does not cause an ignition of the sequence set because it had been inactive and would need two inputs to ignite.
(4) After its ignition, the activated set α reverberates at the highest reverberation level R_1. For some time, its reverberation primes the sequence set that receives strong connections from α continuously. Because there are also strong connections from this sequence set to the other word web, β, that represents a possible successor word B, this set β is primed, too. At this point, one input unit, α, reverberates, and both the sequence detector and the second input unit β are being primed.
(5) B occurs in the input and the primed input unit β therefore ignites. This is now a primed ignition because β had continuously been primed before as a result of α's ignition and reverberation and the strong connections from α to β.
(6) The ignition of β is communicated to the sequence detector. Because it is still being primed by the reverberating α, it responds with an ignition as well.
(7) Finally, ignition spreads back to the reverberating input unit α.

These would be the processes envisaged to occur when the network is presented with a congruent "grammatical" string AB. The important point is that the strong left to right connections in Figure 10.1 guarantee preactivation of the second input unit. This makes its ignition a primed ignition and, therefore, the threshold regulation machinery is not invoked. Ignitions can finally spread backward throughout the activated network, and the primed sequence set also ignites.

Next, the behavior of the network when an "ungrammatical" string BA is being presented is considered.

Network response to an incongruent input BA is as follows:

(1) The word B in the input causes β to ignite.
(2) This is a full ignition $I^\wedge$ without effect on the otherwise inactive network.
(3) The ignition causes temporary priming of the sequence set as a result of backward spreading activity. Assuming that the ignition is a brief

event and backward spreading of activity is faster than the pace of the input elements, this temporary activation vanishes before another input unit is activated.

(4) The word web β's ignition terminates, and this set reverberates at the highest reverberation level R_1. The sequence set and the other input unit, α, are not being primed, because the connections they receive from β are weak.
(5) The word A occurring in the input causes an unprimed full ignition $I^\wedge$ of α. This means that the activity state of α changes from 0 to I.
(6) The strong increase in activity in the net caused by the full ignition activates the threshold control mechanism. The level of reverberation of the already active unit β is therefore reduced from the highest level, R_1, to a lower level, R_2.

After full ignition of α, it is, in this case, not possible that ignitions spread to the sequence detector, because it is not in the primed or reverberating state when α ignites. The three neuronal sets are left at different states of reverberation and priming. Set α reverberates at R_1 and set β at R_2; the sequence detector is primed by α and thus exhibits the highest priming level P_1. No terminal wave of ignitions spreading through the network occurs, and the set specialized for sequence detection fails to ignite.

Note, again, the main differences between the two examples: In the case of the congruent string, there is priming of sequencing and input units representing possible successors. Such priming is not present when the deviant string is processed as a result of the direction of strong connections. Preactivation of the second input unit leads to its primed ignition (instead of full ignition), which does not cause the threshold regulation mechanism to become active. Finally, processing was followed by a wave of ignitions running from β to the sequence set and back to α, leaving all neuronal sets in the highest state R_1 of reverberation. In contrast, the backward wave of activity is not created by the ill-formed string, and the preactivated input units finally reverberate at different levels, R_1 and R_2.

At this point, it appears that a network of neuronal sets designed in this way exhibits three phenomena after a canonical string is presented. These three phenomena do not occur together when an ill-formed string is being processed. The phenomena are functionally and causally related within the given network architecture. Although the phenomena are functionally related, they may be envisaged to be the basis of different cognitive processes.

Phenomenon 1 is that the input units and the sequence detector connected directly to the input units have ignited and therefore reverberate at a high level. More generally, for each input unit activated by an element of the input string, at least one group[2] of sequence sets it is directly connected with is in a high state of reverberation. If one or more sets of sequence detectors connected directly with an input unit are reverberating at high levels of R (either R_1 or R_2)[3], the input unit is called *satisfied*.

Phenomenon 2 is that all reverberating input units the word webs activated by the string in the input are themselves at the highest level R_1 of reverberation. Neuronal sets reverberating at R_1 are called *visible*.

Phenomenon 3 is that all the sets that did ignite in response to parts of the string in the input have finally participated in a continuous wave of ignitions running through the network. This terminal wave of ignitions is initialized by the ignition caused by the last element in the input string. The ignition wave is said to *synchronize* the neuronal sets involved.

Out of the three phenomena – satisfaction, visibility, and synchronization – not a single one is met if the network in Figure 10.1 is presented with the input BA. No synchronization took place as part of the processes caused by the input. One of the input units is not visible: β is left at R_2 rather than R_1. Also, both input units are not satisfied, because the only sequence set they are connected to is primed rather than reverberating at a high level. In contrast, all three conditions are finally met when the well-formed string AB is processed. This simple example illustrates basic mechanisms of sequence detection that are also relevant in the more realistic linguistic examples explained in the Excursuses.

In summary, three criteria can be proposed for acceptance of an input string by a network of neuronal sets:

(i) All input units activated by the string are satisfied
(ii) All of them are visible
(iii) All neuronal sets activated are synchronized

If these criteria of satisfaction, visibility, and synchronization are reached, the network is assumed to "accept" the string and then deactivate the entire representation.

---

[2] In this example, there is only one sequence set. However, in the next section, it is argued that each input unit (representing a word or morpheme) must be connected to one or more sets of sequence detectors. These sets are supposed to be essential for disambiguating string elements in the input.

[3] The motivation for allowing R_2 in sequence detectors here becomes more obvious in following sections in which the processing of lexically ambiguous words is addressed. Briefly, if a word is used twice and as member of different lexical categories, two sets of sequence detectors are assumed to be reverberating, but only one of them is allowed to reverberate at R_1.

10.6 Lexical Categories Represented in Neuronal Sets

10.6.1 Why Lexical Categories?

The primary task of a syntax network is to show a particular activity pattern when input elements occur in a well-formed or grammatical string. The least that the neuronal elements detecting a string AB must achieve is finding out that Propositions (i)–(iii) are correct. (Recall that the activation of a neuronal unit can be described as the verification of the statement that the entity the neuronal element represents was present in the input of the neuronal unit [cf. Section 6.1; McCulloch & Pitts, 1943; Kleene, 1956].)

(i) A occurs
(ii) B occurs
(iii) A occurs before B

Verification of Statements (i) and (ii) can be achieved by *input units*, the neuronal sets that respond specifically if given morphemes, words or meaningful word parts, are heard or read. As detailed earlier, a complex word, such as *switches*, can be broken down into meaningful units, in this case *switch* and the suffix *-es*, that can realize the plural morpheme or the third person singular present verb suffix. Therefore, input units are assumed to respond specifically to either words or morphemes. If input units are conceptualized as neuronal sets, it is clear how the knowledge that a given word has occurred is kept in memory. The input unit α activated by a word or morpheme A in the input verifies and stores the piece of information that A occurred.

To make it possible for the network to decide (iii), whether A occurred before B, it appears straightforward to assume neuronal elements that specialize in sequence detection, such as the sequence detectors or sets proposed earlier. The activity of a neuronal sequence detector would then mean verification of Statement (iii), that A occurred before B, by the network.

One may therefore postulate sequence sets for word or morpheme sequences. Examples would be neuronal sets responding specifically to the morpheme sequences "the machine" or "switch . . . on." However, if each possible word sequence or word pair sequence of a language were to be represented by specific sequence sets, their number would be astronomically high. An obvious solution to overcome this problem is suggested by all grammar theories. The solution is to categorize words into lexical categories and base serial-order algorithms on these grammatically defined word categories (see Chapter 7). The members of one of these categories can replace each other in particular sentence contexts without changing the grammatical status of the sentence. If lexical elements, words, and morphemes, are grouped into 100 or so lexical categories, the number of sequence detectors necessary would be greatly reduced, because sequence sets could connect to a few category representations instead of an extremely large number of pairs of input units. This strategy may therefore offer a less costly solution to the problem of sequence detection (see Chapter 9).

Unfortunately, however, this solution implies that the basic task of deciding whether an input sequence AB is congruent becomes more complicated than indicated at the beginning of this section. It would now consist of verifying the following statements:

(i) A occurred
(ii) B occurred
(iii) A belongs to lexical category a
(iv) B belongs to lexical category b
(v) The element of class a occurred before the element belonging to class b.

Statements (iii) and (iv) must still be spelled out in terms of neurons.

10.6.2 Lexical Ambiguity

The categorization of a word or morpheme into a lexical category may be achieved by links connecting the neuronal set representing the lexical item with neurons representing lexical categories. When a word is perceived, an input unit specializing in detecting this particular word ignites. It then activates a strongly connected set that also receives strong connections from sets representing other words of the same lexical category as the word that occurred in the input. In this case, categorization of the perceived word would be straightforward.

However, the issue of assigning lexical categories to word forms is complicated by the fact that many words can be members of different lexical categories. Here, it becomes necessary to distinguish between a word form or morpheme at the surface and the surface element classified as a particular type of lexical and syntactic unit. A word such as *beat* can be categorized as a noun or as a verb (*the beat* vs. *to beat*). Words of this kind, that is, word forms that can belong to different lexical categories, are called *lexically ambiguous*. A closer look shows that there are different homophonous verbs *to beat* that differ in their *valence*, that is, the number of complements they require. The issue of lexical ambiguity and that of homophonous verbs differing in their valence is addressed in detail in Chapter 7 in the context of dependency grammar.

Following are more example sentences in which the word form *beat* is used as member of different lexical categories:

(1) The beat resonates.
(2) The heart beats.
(3) Betty beats Hans.
(4) Betty beats Hans with a stick.
(5) Betty beats Hans up.

In (1), *beats* is used as a noun; in (2)–(5) *beats* is used as a verb. In (2), only one complement noun phrase is required, whereas another verb *beat*, sharing the word form with the item in (1) and (2), appears to require two complements, and in (4) even three. Finally, the verb *beat* in (5) takes a verb particle as its complement, in addition to two nouns. The different serial-order constraints of the homophonous words have their counterpart in different meanings. The word can be paraphrased by *rhythm* in (1), *pulsates* in (2), *defeats* in (3), *hits* in (4), and *hurts* in (5).

The serial-order constraints of these different but homophonous words can be characterized using syntactic rules, such as the dependency rules (6)–(10).

(6) N (Det /*/)
(7) V1 (N1 /*/)
(8) V14 (N1 /*/ N4)
(9) V145 (N1 /*/ N4, Prep)
(10) V14p (N1 /*/ N4, Vpar)

Dependency rule (6) means that the word used as a noun requires one complement, a determiner to its left. Rule (7) says that the one-place verb requires a noun in the nominative case to its left, and Rule (8) specifies an accusative noun as an additional complement of a different verb type. In (9) and (10), two more verb types are defined with additional complements to their right, a preposition or a verb particle. (For further explanation of the abbreviations used and details about syntactic algorithms, see Chapter 7).

Evidently, the word form *beat* is not the only one to which one of the Rules (6)–(10) apply. Each of these rules operates on a large subcategory of nouns and verbs.

This consideration makes it obvious that assigning a word A to a lexical category a is closely tied to specifying which kinds of words $B_1, B_2, B_3, \ldots$ are required to occur – and therefore usually occur – together with word A.

Extending this idea from words to morphemes, one can state that each verb stem not only requires noun complements, but that it also requires a verb suffix Vs, which also depends on the case and number of the subject noun. The verb subcategory definitions would therefore be as exemplified in the formulas (11)–(14).

(11) V1 (N1 /*/ Vs)
(12) V14 (N1 /*/ Vs, N4)
(13) V145 (N1 /*/ Vs, N4, Prep)
(14) V14p (N1 /*/ Vs, N4, Vp)

Verb complements such as affixes, nouns, and verb particles follow verbs, although at different distances. Similar statements can be made for nouns and other lexical categories. A nominative noun, N1, usually occurs after a determiner, Det, and before the verb, whereas an accusative noun is usually placed after an element of certain verb subcategories, as summarized by (15) and (16).

(15) N1 (Det /*/ V)
(16) N4 (V, Det /*/)

There is, in principle, no limit for the level of detail one introduces in lexical categorization. Distinctions can be made between different noun types according to their grammatical case – nouns in the cases of nominative N1, genitive N2, dative N3, and accusative N4 – or ordinary nouns, N, from proper names, Prop, and pronouns, Pro. More fine-grained lexical categorization can also be motivated by the fact that the subcategories (e.g., nouns, proper names, and prepositions) are subject to different serial-order restrictions. More fine-grained differentiation of lexical categories can also involve semantic criteria, distinguishing, for example, between the objects that can be the subject or object of beating. Clearly, more fine-grained distinctions increase the number of to be distinguished lexical categories. Current dependency grammars propose approximately 100 different lexical categories and subcategories.

The contrast between common nouns and proper names may be used as a final example of a more fine-grained distinction of lexical categories. Articles occur before common nouns but usually not before proper names, Prop. The proper names used in the nominative and accusative cases may therefore be characterized by (17) or (18). Standard nouns have been characterized by (15) or (16).

(17) Prop1 (/*/ V)
(18) Prop4 (V /*/)

This formulation is different from usual formulations in dependency grammars. For formal reasons, grammar theories define the dependency relation as asymmetrical. If A is the dependent constituent and B is the "governing" constituent, the opposite is not possible. If the noun is the dependent category and the verb governs it, the reverse is not the case. However, comparison of the examples used here [e.g., (11) and (15)] shows that this is not so in the present formulation. The dependency relation is assumed to be based on reciprocal connections between word webs and sequence sets. For this reason, it is defined as symmetrical in the present context. The noun and the verb would therefore be considered to be interdependent, and the same would apply for the verb and its suffix. One may want to speak about mutual dependency here. If two lexical categories a and b are interdependent, the dependency relation can be expressed by two rule types, (19) and (20).

(19) a (/*/ b)
(20) b (a /*/)

The idea that lexical categories come in one canonical order is, of course, a simplifying assumption. For some of the serial-order regularities used here, exceptions can easily be found (e.g., "Off went her hat"). Additional detail must be introduced to describe these, and this would be possible by specifying additional regularities by further algorithms.

10.6.3 Lexical Categories as Sets of Sequence Sets

Because lexical categorization of a particular word is so closely tied to the question about which lexical categories occur regularly in the vicinity of the element in question, it may be fruitful to answer the question of lexical categorization in terms of sequence regularities. For this purpose, a sequence set can be conceptualized as a neuronal set specialized in detecting a sequence feature. A sequence feature is a pair of elements of categories a and b in the order "first a then b." A sequence feature would be present in an input string regardless of how long the lag is between the occurrence of the member of category a and that of the member of category b. The sequence detector ignites only if a given sequence feature is present in an input string. After its ignition, the sequence detector stays in a state of reverberation, thereby storing the sequence feature in active memory.

A lexical category representation can now be defined as the union of several sequence detectors, as a set of sequence sets. Examples would be as follows: The category of a nominative noun would be represented by two sequence detectors, one detecting that the element in question followed an article and the other examining whether it is followed by a verb. Accusative nouns may be represented by the union of sequence detectors, one of which is sensitive to a transitive verb of a certain kind preceding the element in question, and the other detecting a preceding determiner. The representation of the category of a transitive particle verb may include sets sensitive to the word followed by a verb suffix, an accusative noun, and a particle. Essentially, for each lexical category label occurring in a bracket of one of the postulated dependency rules, a corresponding sequence detector can be postulated at the neuronal level.

Some of the postulated sequence sets connected to a particular word representation would, so to speak, "look to the past" – that is, be sensitive to particular inputs preceding the word, whereas others would "look forward into the future" – that is, be sensitive to the word followed by other elements. Thus, a lexical category a could be represented by both "future-oriented" or "forward-looking" sequence features, a_{f1}, $\mathrm{a}_{f2}, \ldots,$ a_{fn}, and by "past-oriented" or "backward-looking" features, a_{p1}, $\mathrm{a}_{p2}, \ldots,$ a_{pm}. The neuronal representation of the lexical category would include the corresponding future- and past-related sequence sets, α_{f1}, $\alpha_{f2}, \ldots,$ α_{fn}, and α_{p1}, $\alpha_{p2}, \ldots,$ α_{pm}, respectively.

In the Rules (6)–(20), each lexical category label in the brackets would be analogous to a sequence set. A lexical category label left to an asterisk between slashes corresponds to a backward-looking sequence set, and a lexical category label to its right corresponds to a forward-looking sequence set. It is clear from Rules (6)–(20) that words and lexical categories can be characterized by varying numbers of sequence features, some categories only exhibiting future-oriented features and others only showing regular relationships with their past. This will be further illustrated in Section 10.6.5 in the context of Figures 10.3 and 10.4.

Distinguishing between past- and future-oriented sequence detectors implies a slight modification of the view put forward in Figure 10.1. There, only one sequence detector was introduced for detecting the sequence AB. In the present proposal, this unit would be replaced by two sequence detectors, a future-oriented detector α_f receiving strong input from the input unit α, and a past-oriented detector β_p strongly projecting to the word web β (Fig. 10.2). In addition, strong connections from α_f to β_p and the corresponding weaker backward connections must be postulated. Recall that all connections are assumed to be bidirectional but asymmetrical. Essentially, this modification amounts to a duplication of sequence detectors that does not change basic functional properties of the networks. The modification is introduced to make descriptions of neuronal grammar easier.

A summarizing remark on the symbols used for referring to linguistic and neuronal elements: Word forms and grammatical morphemes are labeled by capital Latin letters, and for lexical categories, lower-case letters are used. Greek letters are used to refer to neuronal sets representing words or morphemes. Sequence sets may be labeled by a lower-case Greek letter (indicating a lexical category the represented feature is part of) plus an index. A “p” or “f” (indicating whether it is past- or future-oriented, respectively) and a number may appear in the index. Sequence detectors can have several labels if they belong to different lexical category representations.

10.6.4 Neuronal Requirements of a Grammar Machine

The following paragraphs present an estimate of the neurobiological requirements of the envisaged machinery of lexical categorization and subcategorization of words in a language.

In principle, the number of past- and future-oriented sequence detectors per lexical category representation may be large, but it is, in fact, difficult to find lexical categories with more than five complement categories. In dependency grammars, verbs, which probably represent the lexical categories with the largest number of complements, are sometimes assumed to have up to four complements (Eisenberg, 1999; Heringer, 1996; Tesnière, 1959). An example would be the verb form “feed” in (21).

(21) Peter feeds Hans soup with a spoon

Counting the verb suffix “s” as an additional complement, the number of complements would be five in this case. Now, each input unit representing a word form or morpheme can be strongly connected to all the sequence detectors included in the representations of its possible lexical categories. Take the extreme case of a five-complement verb, which is lexically ambiguous and can be classified, as a function of context, into one of five different lexical categories, each with another set of five complements. In this case, the input unit of the word form would need connections to 25 sequence detectors. It is not unrealistic to assume that for each word form representation, such a small number of connections are established during language acquisition.

Assuming a language including 100,000 word forms or morphemes, each of which is part, on average, of two out of 100 lexical categories, each defined by five sequence features. To represent the vocabulary of this language, with each word stored by a corresponding neuronal set, 100,000 input units would be required one for each word. As mentioned, free pair-wise recombination of the 100 lexical category labels results in 10,000 possible pairs. However, the grammar network would not need most of these because most of the possible sequences of categories would not occur regularly. To represent

Table 10.2. Attempt at an elementary description of neuronal grammar at different levels. Linguistic entities are listed with their assumed macroneuronal realization in neuronal sets. The number of neurons involved at the microneuronal level is also estimated. According to this estimate, the grammar circuits require surprisingly few neurons. The following assumptions underlie the computations: One neuronal set includes 100,000 neurons. One language includes 100,000 words. The words of a language fall into 100 different lexical categories. Each lexical category is characterized by 10 or less sequence features.

Linguistic	Macroneuronal: sets	Microneuronal: neurons
Word/morpheme	1 input unit each	10^5 neurons
Sequence feature	1 sequencing unit	10^5 neurons
Lexical category	$<$10 sequencing units	$<10^6$ neurons
Lexicon	10^5 input units	10^{10} neurons
Grammar	10^3 sequencing units 10^7 connections between sets	10^8 neurons

the 100 lexical categories, each with a maximum of five complements, 500 sequence detectors would be required. Doubling the number of sequence detectors, as suggested for illustration purposes, would result in a maximum of 1,000 sequence sets. Thus, 1,000 sequence detectors would be required for representing a complete grammar in the way outlined here. The number increases if more lexical categories are being postulated.

Nevertheless, the preliminary result is that the number of word or morpheme representations would by far outnumber the number of sequence detectors necessary. Assuming that the size of a neuronal set is 100,000 neurons, the maximum number of neurons of the vocabulary or lexicon would be 10^{10} and the maximum number of grammar neurons making up sequence detectors would be 10^9. These numbers of neurons are available for language processing in the cortex. Table 10.2 summarizes the present estimate of the "size" of the serial-order machinery. Note again the strikingly low number of neurons required for wiring syntactic information.

The number of axonal connections between input and sequence sets would probably be much larger than the number of neurons involved. For connecting the 100,000 input units to their proposed five category representations, each of which is made up of five sequence detectors, a few million connections linking input and sequence sets would be required. Since most cortical neurons have above 10^4 synapses, the large number of connections should not constitute a problem. Note that not all neurons of a set need to connect to another set to establish a strong functional relationship between both sets. If there are 10^7 macro-level connections between neuronal sets each constituted by 1,000 axons, the number of neuronal links would be 10^{10}, thus only a small percentage of the $\sim 10^{15}$ axons of the neurons involved. Again, the proposal appears realistic.

According to these estimates, a number in the range of 10^5 neuronal sets is necessary for representing the vocabulary or lexicon of a language, whereas the more narrowly defined grammar network, that is, the neuronal ensembles specifically representing serial-order information, would be much smaller. Only if a substantially higher number of lexical categories is postulated (e.g., to 10,000 categories), the number of necessary sequence detectors would reach the number of stored words or morphemes (10^5). Thus, an increase in the estimate of the number of word categories and sequence detectors leads only to a relatively small increase of the size of the required network. The introduction of new sequence detectors responding to additional features of a string opens up the possibility of representing more word categories, including, for example, semantic subclasses of lexical categories, in an economical way.

10.6.5 Lexical Disambiguation by Sequence Sets

A neuronal set representing a word form such as *beat* or *switch*, which, depending on within-sentence context, is part of one of several lexical categories, can now be assumed to be linked closely to all the neuronal sets representing these alternative lexical categories. However, for each of its occurrences, the word form should be classified as belonging to only one of the categories. This requires a mechanism deciding between the neuronal sets representing alternative lexical categories to which a given word can be assigned. The mechanism can be realized by a regulation mechanism, for example, mutual inhibition between the alternative representations of lexical categories. The most active lexical category representation would therefore become fully active and the competitor category representations would be suppressed.

More precise criteria that could allow for categorizing a word form A as either a member of lexical category a or b could be the following: A could be classified as a member of a if all sequence sets included in the representation of the a category have ignited a few time steps earlier, and if there is at least one sequence detector included in the representation of b that did not ignite. This allows only for deciding between sets of sequence detectors, each of which includes specific sequence sets. However, if the representation of one lexical category representation includes the set of feature detectors of

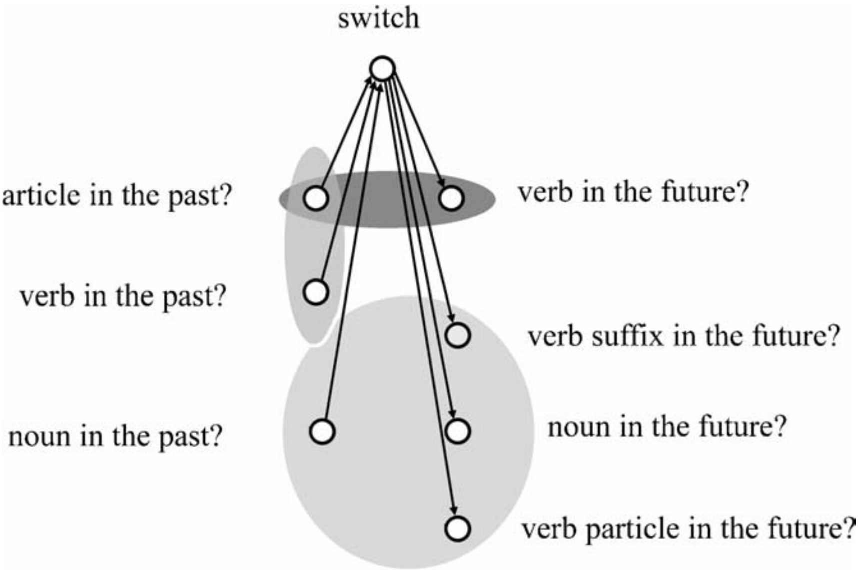

Figure 10.3. Putative neuronal organization of lexical categories. An input unit representing the word form *switch* is connected to three competing lexical category representations, as indicated by ovals in different tones of gray. The three alternative lexical categories are nominative noun, accusative noun, and transitive particle verb. Each category representation consists of two or more sequence sets, indicated by small circles. Each of the sequence sets responds specifically to a different feature of the context, as indicated by questions. Inhibition between lexical category representations is provided by an inhibition mechanism. Modified from Pulvermüller, F. (2000). Syntactic circuits: how does the brain create serial order in sentences? *Brain and Language 71*, 194–9.

the other [as would be the case for V1 and V14, cf. (7) and (8)] the decision criterion must be refined. For example, if the last inactive sequence detector included in the representations of both a and b ignite at the same time, the set with the larger number of sequence detectors wins.

This or a similar "winner-takes-all" dynamics can be implemented by a network of lexical category representations, each composed of one or more sequence detectors if inhibitory connections are present between all representations of lexical categories. Figure 10.3 sketches the putative neuronal wiring of three homophonous words. Figure 10.4 presents the same wiring with the more abstract labels used in Section 10.6.3. In this illustration, one input unit, the neuronal counterpart of the word form, is connected to three sets of sequence sets. These sets are indicated by ovals in different shadings of gray. The ovals are labeled by Greek letters, each of which would be thought to represent one of the lexical categories the homophones can be part of. Each set comprises two to four sequence detectors labeled by

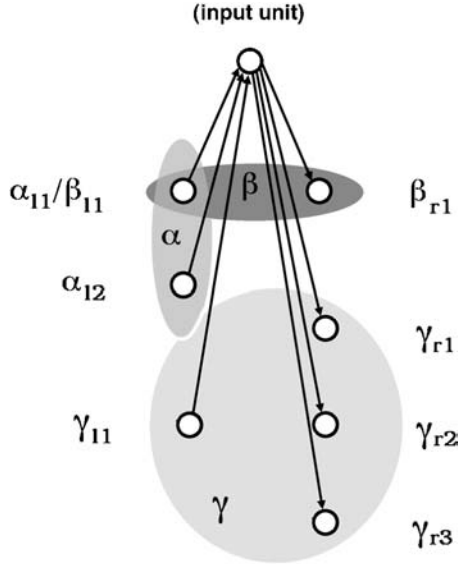

Figure 10.4. Putative neuronal organization of lexical categories: more abstract formulation. An input unit is connected to three competing lexical category representations, α, β, and γ, each consisting of two to four sequence sets labeled by a Greek letter plus index. The sets of sequence sets organizing lexical categories inhibit each other so that only one of them can become fully active together with the input unit. (A more concrete case is discussed in Fig. 10.4.)

Greek letters plus index. To guarantee inhibition between lexical category representations, the sequence detectors would be connected reciprocally to additional neuronal elements that would be connected among each other by inhibitory connections. This inhibitory wiring scheme is reminiscent of the one proposed for corticostriatal circuits (see Fig. 5.2). The inhibitors would be activated only if all the sequence sets included in their respective lexical category representation were active at a time. Furthermore, this type of between-category inhibition would become effective only if a word form activates more than one lexical category representation at a time. If all sequence sets included in a lexical category representation reverberate at the same time, they coactivate the inhibitor that, in turn, feeds excitation back to the sequence sets. The input from the inhibitor would not be assumed to be essential for maintaining enhanced activity levels in sequence detectors [see principles (A3) and (A3') below].

As emphasized, inhibitory interaction between alternative lexical category representations of a word form is needed for disambiguation of lexically ambiguous word forms. Although no data are available that would allow for narrowing down the possibilities for realizing mutual inhibition of lexical representations, a mechanism of this kind is a necessary component of a neuronal grammar. This wiring scheme may have a possible brain basis in corticostriatal connections and inhibition in striatum, but could also be realized by other direct or indirect inhibitory links between cortical representations.

How would such a circuit actually solve the task of categorizing a lexically ambiguous word form as a member of one of its possible lexical categories? The "decision" should be made by a circuit including mutually exclusive sets of sequence detectors, each representing one of the possible lexical categories. To answer this question, consider again the more concrete example of the word form *switch*. Depending on the context in which it occurs, it can be classified, for example, as a noun in the nominative or accusative case, N1 or N4, or as a transitive particle verb, V14par. The characteristics of these lexical categories can be expressed by dependency rules formulated earlier in this section, repeated here as (22)–(24) for convenience.

(22) N1 (Det /*/ V) Nominative noun
(23) N4 (V, Det /*/) Accusative noun
(24) V14par (N1 /*/ Vs, N4, Vpar) Transitive particle verb

This formulation in terms of syntactic algorithms can be translated into the language of neuronal sets, as illustrated by Figure 10.3.

The alternative lexical categories specified by (22)–(24) are represented by sets of sequence sets, as indicated by ovals at different gray levels. Each complement required by one of the three possible lexical categories of the word form has its corresponding sequence detector in the figure. The mutual inhibition of lexical category representations ensures that only one of the three alternative sets of sequence detectors is active with any occurrence of the word form *switch*.

One of the sequence feature detectors, the uppermost on the left in the figure, responds to the word form *switch* that after an article (or determiner) occurred in the past. It ignites if a word form that can be used as an article had occurred and the word form *switch* (or any other word form that can function as a noun) follows it. On the upper right is a sequence set responding to the word *switch* followed by a verb. This sequence detector can be thought of as checking the near future for any word form that can be used as a verb. Because it is being primed by the occurrence of *switch*, it ignites when a verb follows it. If both the past-oriented sequence set responsive to an article and the future-oriented sequence detector targeting a verb have ignited, the complete representation of the nominative noun category (oval in dark gray) reverberates. The word form unit is therefore satisfied and assigned, by the network, to the respective lexical category. This happens in response to a string such as (25).

(25) *The* switch *broke*.

The neuronal representations of the words in italics are assumed to be linked to the word web of *switch*. Similar processes take place if a verb and an article both precede the word form *switch*, as in sentence (26).

(26) Betty *turned* the switch on.

In this case, two past-oriented sequence sets, primed by determiner and verb, respectively, ignite when the word form *switch* occurs.

If the word form *switch* occurs after another word that can be used as a noun, the sequence detector on the lower right ignites. Given the word form is also followed by a verb suffix, still another noun (in the accusative case) and a particle, three more sequence sets sensitive to future events ignite (after being primed by *switch*). These would be processes triggered by the input string (27).

(27) The *man* switched the *engine* off.

In this case, the word form *switch* would ignite in close vicinity to the ignition of all sequence detectors defining the neuronal representation of the lexical category of a transitive particle verb.

In response to the example sentences, sequence detectors from nontarget lexical category representations may also ignite. For example, when sentence (27) is presented to the circuit fragment depicted in Figure 10.3, the sentence-initial article primes the uppermost sequence detector on the left, whereas the second word of the sentence, the noun *man*, primes the sequence set at the lower left. The word from *switch* in the input ignites both primed sequence detectors. Nevertheless, the further evidence accumulated when the additional words of the string (27) are processed activates all sequence detectors of only one lexical category representation, that of the transitive particle verb. Between-category inhibition is assumed to minimize activity in the competing lexical category representation.

It should be mentioned that this simple description glosses over a few facts that may complicate the operation of a mechanism of this type. In the earlier formulation in rules (22)–(24), different noun types marked for nominative or accusative case were distinguished. Making this distinction based on the surface form of words is perhaps impossible in English, because nominative and accusative case are not overtly marked in this language. Furthermore, the verb suffix may also not be realized; for example, in the first person singular present case. In English, lexical subcategorization may primarily rely on serial order information. Further relevant information about lexical subcategories may be extracted from phonetic properties of spoken language. The distinction between grammatical cases is most obvious in languages where these are overly marked, for example by specific inflectional suffixes.

After the input unit of a word has ignited and therefore reverberates, it may be possible that a large number of future-oriented sequence sets, which are part of representations of possible lexical categories of the word, are primed. This activity is transmitted to all of the sets representing words of lexical categories that can be successors of this word. A rather large number of input units are therefore primed. Nevertheless, a successor word selects and fully ignites only one of the sequence feature detectors connecting its representation to that of *switch*. If only one of the possible successor words whose input units are primed occur in the input, the respective input unit exhibits primed ignition and ignitions spread back to the input unit of *switch*. In this case, the input unit is synchronized with (part of) the verb representation. This happens until all sequence sets of the verb representation have been ignited and therefore reverberate. Inhibition between lexical category representation finally suppresses the competing category representations.

These examples illustrate that in a network of neuronal sets, both forward and backward flow of activity can contribute to disambiguation and classification of lexically ambiguous words into one of the lexical categories they can be part of. The classification can be complicated; for example, if a word exhibits multiple possible relationships with many other elements of a string, or if two word forms with the same lexicosyntactic properties occur in the same sentence. Some of these problems are addressed in the more formal reformulation of the proposal in the next section and in Chapters 11 and 12.

The general idea put forward in this section is that the possible classifications of a first element A in a string allow to "nominate" (by priming) potential successors $B_1, B_2, \ldots$, which, if one of them is chosen (i.e., ignites), in turn contributes to disambiguation of the initial element A, thus "approving" one of its possible context-sensitive classifications. The idea is very simple and is motivated by neuroscientific considerations. It appears necessary and may be fruitful to spell out its possible brain basis in more detail. The mutual nomination and approval of classifications of ambiguous elements occurring in a string requires reciprocal connections between their context-sensitive representations.

Note that this idea has some relationship to what Dilthey (1989), a philosopher, called the *hermeneutic circle* – that is, the development of a higher-order understanding of a symbol, or chain of symbols, on the basis of the processing of a second symbol whose interpretation is, in turn, influenced by the earlier symbols. In the present illustrations, only crude lexicosyntactic classifications (noun or verb) and subclassifications (one- or two-place verb, nominative or accusative noun) were made, but, of course, they have semantic implications. The idea may well be extended to more fine-grained semantic distinctions. Disambiguation of an earlier element is achieved by the occurrence of (a set of) later elements with which it has a regular serial-order relationship, and conversely, the later element(s) can also be classified based on information provided by the earlier element. For this to operate, activity caused by symbols in the input must be assumed to reach at least the representation of possible successors of the symbol and must be allowed to flow back from the successors' representations to the representation of the original one. This mechanism can be spelled out in the language of neuronal sets.

10.7 Summary: Principles of Neuronal Grammar

For making scientific progress, theorizing is necessary, and a newly developed model needs to be illustrated in a concrete manner. To define functional principles in a way that would make computer simulation and animations possible, some of the assumptions must be made more precise. To this end, arbitrary decisions are sometimes necessary. To focus on a concrete example, there is no strong reason to assume, as proposed here, that activity of a particular input unit only primes input units of its possible direct successors. A two-step or higher-order priming process would also be possible, although computationally more demanding. As a second example, it is necessary to restrict activity in the network, but there are several options for choosing the exact criterion for activating the threshold control mechanism.

Recall that it is the primary purpose of the present considerations to demonstrate syntactic competence in neuronal networks that may be housed in the brain. (The fact that they could be organized slightly differently is irrelevant for this purpose.) Consideration of alternative possible neuronal mechanisms is, of course, highly relevant. It is possible, or even likely, that details of the networks proposed here are incorrect as descriptions of actual circuits in the brain. They can be modified once the necessary experiments have been performed.

A few terminological remarks are necessary before the principles assumed to underlie the dynamics of the proposed circuits can be formulated. The principles may reflect properties immanent to the human central nervous system. Without making this assumption, one may prefer to consider them axioms of the grammar networks proposed. Six axioms, (A1)–(A6), are formulated, some of which will be reformulated later. The predicate $S(x, y)$ specifies activity state x at time step y of neuronal set S. Although there are, in principle, infinitely many different activity states (see Section 10.2), only four types are distinguished: rest (0), ignition (I), reverberation (R), and priming (P). As described in Section 10.2, different levels of reverberation are called R_1, R_2, R_3 (or R, R', R'') and so on, with indices giving the rank in the activity hierarchy. Larger numbers therefore mean lower activity levels. Priming levels are labeled in the same manner, as P_1, P_2, P_3 (or P, P', P'').

Now, a basic set of assumptions is as follows: The initial activity state of all sets is rest, 0. If there is no input to an inactive neuronal set, it stays in state 0. If a set is reverberating, it stays at its reverberation state R_i– although, as discussed in Section 10.2, its absolute level of activity may fall off with time. Ignition, I, and priming, P, are short-lived events. I is immediately followed by reverberation (R_1), and priming by inactivity (0). After an ignition, there is a period of reduced excitability preventing too-frequent ignitions. The double arrow ($\Rightarrow$) means "causes and is followed by."

These assumptions can be expressed by (A1).

(A1) Spontaneous changes of activity in set S

- (i) $S(0, t) \Rightarrow S(0, t + 1)$
- (ii) $S(I, t) \Rightarrow S(R_1, t + 1)$
- (iii) $S(P_i, t) \Rightarrow S(0, t + 1)$
- (iv) $S(R_i, t) \Rightarrow S(R_i, t + 1)$
- (v) Refractoriness: If $S(I, t)$, then S cannot ignite at $t + 1$ or at $t + 2$.[4]

The effect of priming at time t has vanished one time step later, at $t + 1$. Therefore, input at t to a set exhibiting state P but receiving no additional input at t essentially has the same effect at $t + 1$ as input to a set at rest.

Spontaneous activity changes as summarized in (A1) are one factor determining network dynamics. There are four possible causes of activity states in neuronal sets that will be considered.

- (a) External input to input units.
- (b) Activity flow between sequence sets and input units.
- (c) Interaction between neuronal sets and threshold regulation.
- (d) Inhibition between sets in the case of processing of ambiguous words.

Input to the network from outside by an element of an input string is called *external stimulation*, E. External stimulation of an input unit causes it to ignite and reverberate later. The primed ignition I of an already highly active unit contrasts with the stronge full ignition $I^\wedge$ of a previously inactive unit. S (E, y) denotes the input E at time y to set S.

---

[4] This principle overrides the principles formulated by (A3) and (A4):Immediately after an ignition, a set cannot go back to I.

(A2) Activity change in an input unit S caused by external stimulation

(i) $S(0, t),\ S(E, t) \Rightarrow S(I^\wedge, t+1)$

(ii) $S(P_1, t),\ S(E, t) \Rightarrow S(I, t+1)$

(iii) $S(P_i, t),\ S(E, t) \Rightarrow S(I^\wedge, t+1)$
if $i > 1$

(iv) $S(R_1, t),\ S(E, t) \Rightarrow S(I, t+1)$

(v) $S(R_i, t),\ S(E, t) \Rightarrow S(I^\wedge, t+1)$
if $i > 1$

The difference between full ignition $I^\wedge$ – as specified by (A2) (i), (iii), and (v) – which activates threshold regulation [see (A6)] and primed ignition I [cf. (ii) and (iv)] which has no additional global effect on network dynamics, is important.

Clearly, if priming or reverberation levels are very low, their effect becomes negligible. Therefore, the principles addressing priming and reverberation effects on ignition must be specified with regard to the activity levels for which they apply. Regarding principle (A2), this cut-off is assumed to be below P_1 and R_1 – that is, only the highest levels of priming or reverberation contribute to neuronal processing in the network as specified by (A2). Also, slightly lower levels of activity (P_2 and R_2) are assumed to play a role in the interaction of neuronal elements as specified by (A3) and (A4). Levels of priming and reverberation below R_2 and P_2 (and sets exhibiting only such low levels) are called *invisible*.

Both input and sequence sets, receive internal stimulation – that is, excitation through connections from other sets. A long arrow ($\longrightarrow$) is used to indicate a bidirectional link with stronger connections in the direction of the arrowhead. Recall that direction and strength of connections between input and sequence sets carry information about the input strings that are accepted by the network.

Ignition of a set spreads through all connections to sets already exhibiting a high activity level (i.e., reverberation or priming at rank 1 or 2). If S_q is less active or at rest when set S_p ignites or reverberates, activity of S_p primes S_q only if the connection from S_p to S_q is strong. In this case, priming affects not only strongly and directly connected sequence sets, but priming is further communicated to the next input unit also.

(A3) Activity changes caused in S_q through a connection between S_p and S_q

(i) $S_p(I, t),\ S_q(R_i, t) \Rightarrow S_q(I, t+1)$
only if $i = 1$ or $i = 2$

(ii) $S_p(I, t),\ S_q(0, t) \Rightarrow S_q(P_1, t+1)$
only if $S_p \longrightarrow S_q$

(iii) $S_p(R_i, t), S_q(0, t) \Rightarrow S_q(P_i, t+1)$
only if $S_p \longrightarrow S_q$

(iv) $S_p(P_i, t), S_q(0, t) \Rightarrow S_q(P_i, t+1)$
only if $S_p \longrightarrow S_q$ and S_p is a sequence set

Spreading of activity through connected sets needs time. If external stimulation reaches an input unit, it ignites one time step later, and during the following time steps, priming is exerted on its future-oriented sequence set, which in turn prime connected past-oriented units projecting onto further input units. Thus, there is a lag of four time steps between the occurrence of a symbol in the input and priming of the representations of its possible successors in the string. Only if three or more time steps separate the occurrence of two consecutive symbols in the input is the priming machinery effective.

To pronounce a word consisting of one syllable, it takes approximately 150 ms or more although some inflectional affixes may be realized as even shorter events. To allow for three time steps between the activity waves caused by two one-syllabic words that directly follow each other, the "time steps" assumed for the present syntactic derivations need to correspond to 50 ms or less of real time, corresponding to a frequency of 20 Hz or higher. This applies, for example, to the spreading of priming from one input unit to its sequence set and to additional input units, or to the spreading of ignitions from one set to a neighboring primed set. There are also longer-lasting activity processes in the present neuronal grammar machinery. For example, the backward spreading wave of ignitions representing an entire sentence would be a much longer-lasting process, and importantly, the reverberation of a set is assumed to last at least for tens of seconds.

In the derivations shown next, it takes up to seven time steps after activation of an input unit by external stimulation until all relevant computations have been performed. Only then will the system have reached a steady-state in which spontaneous activity changes have come to a rest. To avoid effects that are related to the exact timing of incoming words – pauses or slight variation in the speed of speaking not considered here – the next input from outside should reach the network when it has settled into a stable state, that is, only after seven time steps. Thus, to be on the safe side, it may be best to conceptualize the time steps in the present computations as even shorter intervals, of some 20–30 ms of real time in real networks. This corresponds to a frequency of $\sim$30–50 Hz.

One remark relevant to subsequent example simulations is necessary here: If there are several past-oriented sequence sets in a lexical category representation, then (A3iii) and (A3iv) must be modified as follows: Priming of an input unit of a possible member of such a lexical category is achieved only if all the past-oriented sequence sets of the lexical category representations are primed. The level of priming is calculated as the average of activity levels exhibited by the sets exerting priming on the input unit. If the average is not a natural number, the next smaller number is chosen.

More complicated issues in the activation dynamics of the proposed syntactic circuits must be addressed. Imagine that two sets feed into a third one. How would their activity states interact with the recipient to yield new states?

If a given set S_r is inactive but receives excitatory input from two other sets S_p and S_q, simultaneous input through both connections may summate. To change the state of a sequence set from inactivity to reverberation, at least two simultaneous inputs are necessary. One input caused by an ignition is necessary and the other input may also be ignition related, or may be mediated by reverberation or priming communicated through strong connections.

(A4) Activity changes caused in S_r through connections with S_p and S_q

(i) $S_p(I, t), S_q(I, t), S_r(0, t) \Rightarrow S_r(I, t+1)$

(ii) $S_p(I, t), S_q(R_i, t), S_r(0, t) \Rightarrow S_r(I, t+1)$
only if $i = 1$ or $i = 2$, and $S_q \longrightarrow S_r$

(iii) $S_p(I, t), S_q(P_i, t), S_r(0, t) \Rightarrow S_r(I, t+1)$
only if $i = 1$ or $i = 2$, and $S_q \longrightarrow S_r$, and S_q is a sequence set

Thus, two simultaneous excitatory inputs, one of which must be an ignition, causes a given set to be "switched on" – that is, to ignite and reverberate later. In this case, R and P again exert their influence only through strong connections, and only if their activity rank is high (R_1 or R_2; P_1 or P_2).

Only if ignitions in two sets are exactly synchronous can they cause an additional ignition in a connected third set according to principle (A4i). Because the duration of words in spoken sentences and the delays between them (and the scanning times for individual words during reading) can vary substantially without affecting sentence comprehension, it would not be wise to base the network's grammaticality judgments on variations of timing of neuronal activation processes caused by the speed of the language input. Therefore, only axioms (A4ii) and (A4iii) are relevant in the following derivations. As mentioned, short time steps are assumed for the computations in the network so that the network settles into a stable state by the time a new input unit is activated.

Global network dynamics may exert an additional important influence on the computations performed by syntactic circuits. Threshold regulation is activated by full ignition [see axiom (A2i)] and leads to reduction of all levels of reverberation from R_i to R_{i+1}. Additional activations of threshold regulation reduce the level to R_{i+2}, R_{i+3}, and so on. Levels of priming are affected in the same way. The regulation mechanism may also enhance neuronal activity if its level is low. In the present formulation, reverberation and priming levels are just relabeled so that the highest levels are always R_1 or P_1. Nevertheless, lower activity levels are not adjusted so that not all levels between the maximal and minimal level are necessarily present at all times (e.g., the only states may be R_1 and R_3).

(A5) Threshold regulation

- (i) if a full ignition $I^\wedge$ of an input unit happens at t, then for all reverberating S:
  $S\,(R_i, t) \Rightarrow S\,(R_{i+1}, t+1)$
- (ii) if there is no ignition or reverberation at R_1 at t, then - for all reverberating S:
  $S\,(R_i, t) \Rightarrow S\,(R_{i-1}, t+1)$

Because priming is caused by reverberation [as specified by (A3iii)] and by ignition (A5), implies parallel adjustment of priming levels.

As detailed in Section 10.6 (the section on lexical representations), alternative lexical categories of a given word or morpheme must exclude each other. The postulated inhibition mechanism is assumed to exclude simultaneous ignition of two lexical category representations connected with one input unit. Also, if part of a lexical category representation ignites or reverberates at R_1, activity of competing lexical category representations connected with the same input unit are assumed to be reduced. The respective input unit must be active (at I or R_1) and thereby enable category inhibition.

(A6) Inhibition between two lexical category representations α and β connected to an active input unit S that ignites or reverberates at R_1:

- (i) Ignition of S can only spread to sets included in either α or β. The most strongly activated representation wins.
- (ii) If one or more sets included in α but not in β ignite at t, then no set included in β and not included in α can exhibit I, P_1, or R_1 at $t+1$.

(A6i) is formulated without giving an exact criterion for calculating activity levels of lexical category representations, because it is only occasionally relevant in the circuits here (cf. Section 10.5.4). To decide which already active alternative category representation, α or β, ignites, the average activity level of the active sets in α and β can be chosen as an additional criterion. If activity strength is calculated by averaging, three sets at R_1, P_1, and R_2 have a higher average level (1.3) than two sets at R_1 and R_2 (average: 1.5).

If both competing category representations exhibit the same average level of preactivity, ignition spreads to the larger set (the one with the larger number of sequence sets).

In Chapter 12, the principles formulated by (A2)–(A6) are modified to improve the present version of a neuronal grammar.

It is important that principles (A1)–(A6) are thought to reflect biological properties of the nervous system described at the level of neuronal ensembles and their interactions. Some of them are closely related to the established neuroscientific principles discussed in Chapter 2 or to the mechanisms outlined in Chapters 5, whereas others are based primarily on brain-theoretical postulates (see earlier discussion). Principles (A1)–(A6) can be viewed as high-level functional consequences of neurophysiological properties of neurons and the way they are connected neuroanatomically in the human brain. If this view is correct, the principles reflect neuronal structure and function as determined by the human genome.

CHAPTER ELEVEN

Neuronal Grammar and Algorithms

This chapter addresses the question of how to translate grammatical algorithms into the language of neuronal sets.

11.1 Regular Associations, Associative Rules

There has been some discussion about the question of whether the human mind and brain use neuronal principles and connections for processing grammatically related information, or whether it uses rules and algorithms specified by grammar theories (Elman, Bates, Johnson, Karmiloff-Smith, Parisi, & Plunkett, 1996; Pinker, 1994). This proposal suggests that these positions, which are sometimes considered to exclude each other, are, in fact, both correct. This is not meant in the sense that there are two modules or systems, one for neural networks and the other one for rule algorithms (Pinker, 1997), but in the sense that rules are abstract descriptions of the neuronal machinery, as they are, without any doubt, descriptions of aspects of human behavior and action (Baker & Hacker, 1984).

If rules and algorithms are adequate descriptions of aspects of human behavior and action, they must have a basis in neuronal structure and function. As stressed in the discussion of the McCulloch–Pitts theory (Section 6.1), a neuronal network can be reformulated using calculus or by a logical formula. It is therefore reasonable to ask which putative neurobiological counterparts exist for syntactic rules and, conversely, how a neuron circuit sensitive to serial order can be adequately described algorithmically.

The cortex is an associative memory and would therefore be difficult to imagine it ignoring the correlation of words and morphemes in its input. Rather, it may use the information about correlation of word pairs for setting up sequence sets and their connections to the representations of word groups whose members can be replaced by each other in a given context.

The neuronal grammar wired up in this manner may finally exhibit properties very similar to aspects of rule descriptions in grammar theories. One type of grammar model, dependency syntax, can, after some modification, be translated straightforwardly into the language of neurons, as has earlier been suggested and is elaborated further in this chapter. Neuronal grammar may be rooted in and limited by general principles immanent in neuronal function.

From one viewpoint, each individual neuron performs a logical computation (Section 6.1), and therefore greater neuron ensembles would do so as well. Strongly connected neuron ensembles are particularly well suited to represent the elements of logical or algorithmic operations, because they exhibit discrete activity states. The functional web or neuronal set either ignites or does not. This functional discreteness can be assumed to be present, although in different contexts of background activity, the exact boundaries of the neuron set may slightly vary.

The equivalence of algorithmic descriptions and neuronal circuits can be illustrated by spelling out linguistic rules and regularities in the language of artificial neurons. For example, the rules of regular past-tense formation and the regularities underlying past-tense formation of irregular verbs can be contrasted using a neuronal automaton, and individual connections and connection bundles can be correlated with the storage of information about rules and exceptions (see Chapter 6; Pulvermüller, 1998). This chapter further illustrates the equivalence of algorithmic descriptions and networks by providing fragments of neuronal grammar depicted as neuronal circuits, and also as algorithms.

The close relationship between a description in terms of syntactic algorithms and a description of properties of syntactic circuits is also evident from the summary of Chapter 10. There, principles underlying dynamics of neuronal grammar were presented in the form of axioms (A1)–(A6). A reader who prefers algorithms to neuronal networks is free to take these axioms to express assumptions simply that allow for economical modeling of abstract syntactic processes. From a neurobiological perspective, however, these same axioms are, in the best case, correct descriptions of principles of neuronal dynamics in the human brain. In this case, the algorithmic and neuronal formulations would be near equivalent.

Axioms (A1)–(A6) may formulate what all grammar circuits have in common: their putative common ground in properties of the neurobiological machinery. In this chapter, the focus is on the specific adjustments necessary for representing a particular language. To speak a language, one must not only have a largely intact brain, but must one also have learned words, their assignment to lexical categories and subcategories, and the restrictions that apply to ordering them in time. Clearly, axioms (A1)–(A6) have implications for the possible temporal ordering, but they also allow for some language-specific variation. The assumption is that the latter can be learned on the basis of Hebbian associative learning principles.

11.2 A Formalism for Grammar Networks

As already emphasized, syntactic circuits in the human brain, the neuronal side of grammar, can be characterized by

- neuronal and grammatical principles (see Section 10.)
- formulas representing language-specific regularities and wiring.

This section shows how the formulation can be adjusted so that for each formal statement of the second type, there is a neuronal circuit or connection pattern to which it refers. The assumption is that there are 1 : 1 correspondences between linguistic representations and neuronal entities (e.g., neuronal sets) and between linguistic processes and neuronal processes (Pulvermüller, 1992).

As mentioned, the formulas representing linguistic regularities (and wiring schemes) use symbols for *terminal elements*, which refer to word forms and affixes actually occurring in sentences (cf. Chapter 7). In addition, nonterminal symbols for lexical categories (referring to word and affix types) are used. As detailed, words and morphemes, the terminal elements, have their putative organic basis in word webs or input units, whereas lexical categories, one type of nonterminal symbol, are realized as sets of sequence sets – "sequence supersets," so to speak.

What follows now is one possible statement of neuronal grammar in terms of algorithms. Three types of formulas are specified: assignment, valence, and sequence formulas.

Assignment formulas specify the correspondence between word forms or affixes and lexical categories. This many-many assignment function (Bar-Hillel, Perles, & Shamir, 1961) is specified by a set of assignment formulas, each giving one or more lexical elements and the lexical category to which they belong. A double arrow (↔) is used to denote the assignment relation.

(1) $\mathsf{A, B, C, \ldots \leftrightarrow a}$

By (1), the lexical elements A, B, C, and so on are assigned to lexical category a. Each lexical element, word, or morpheme can be assigned to more than one lexical category.

Assignment formulas are proposed to be organized in the brain by reciprocal but asymmetrical connections between input units and a collection of sequence sets.

Valence formulas, included in a language-specific neuronal grammar, specify properties of a lexical category in terms of sequence features.

(2) $a\ (p_1, p_2, \ldots, p_m /{*}/ f_1, f_2, \ldots, f_n)$

By (2), lexical category a is characterized as having *m* backward sequence features (briefly, *backward features*) specifying categories regularly occurring within a sentence before an element of category a occurs. The *n* forward sequence features (briefly, *forward features*) specify what must be expected in the future given an element in the input is classified as member of lexical category a. The asterisk between slashes indicates the relative position of the element assigned to a. This is similar to giving the valence of a lexical category and listing how many of the complements must occur to the left and right of a member of the category. Therefore, (2) is called a valence formula. Note that no order of the preceding elements p_i is implied, as there is no preferred order of the elements f_j following the member of the category a.

The neuronal substrate described by valence formulas would be the backward- and forward-oriented sequence sets connected to input units of a certain kind. These input units frequently share their neighbors in sequences of ignitions. They represent members from a given lexical category.

Sequence formulas specify individual sequence features and have the following form:

(3) $a\ (f_i) \longrightarrow b\ (p_j)$

The long arrow ($\longrightarrow$) can be read "connects to." In neuronal grammar, this type of connection provides an important mechanism of serial ordering. Sequence formulas have their putative neuronal equivalent in reciprocal but asymmetric connections between sequence sets.

The following two sequence formulas would be relevant for storing the knowledge that certain types of English nouns and verbs follow each other [see (4)], and that certain types of English verbs are followed by a particle [see (5)].

(4) $N\ (f) \longrightarrow V\ (p)$

(5) $V\ (f) \longrightarrow \text{Vpart}\ (p)$

Sequence formulas can be inserted into valence formulas to yield a description of lexical categories in terms of which other categories are required to the left and right. For example, inserting formulas (4) and (5) into (6) yields formula (7).

(6) V (*p* /∗/ *f*)
(7) V (N /∗/ Vpart)

Formula (7) implies that an element categorized as an English verb (V) of a particular subtype can come with a preceding noun (N) and another element following it that can be categorized as a verb particle (Vpart). Clearly, the verb category meant here is an intransitive particle verb (V1part). Derivations of sequence and valence formulas such as (7) are called *rules of neuronal grammar* and resemble valence rules used in dependency grammars (see Chapter 10 for discussion of common features and differences).

Assignment formulas can also be inserted into valence formulas or into rules. By inserting the assignment formula (8) into Rule (7), the word-specific Rule (9) can be obtained.

(8) get ↔ V
(9) get (N /∗/ Vpart)

Rules of neuronal grammar provide descriptions of sentence structures. Such structural descriptions are obtained by inserting valence formulas or rules into each other. The following set of abstract rules provides an example:

(10) c (b /∗/ e)
(11) b (a /∗/)
(12) e (d /∗/)

Inserting (11) and (12) into (10) yields bracketed complex rule (13).

(13) c [b (a /∗/) /∗/ e (d /∗/)]

The implications of (13) are the same as those of (11) and (12) applied after (10).

11.3 Some Differences Between Abstract and Neuronal Grammar

These examples of insertions of formulas into formulas may remind one of what is possible with standard grammar algorithms. One may therefore want to state that conventional grammars, for example, context-free rewriting systems or dependency grammars (see Chapter 7) can be reformulated by a language-specific version of neuronal grammar. However, this would be incorrect for the following reasons.

First, a rewriting system allows for multiple use of the same rule and lexical category representation. A rule can be used recursively in conventional grammar models. This is not so in the present neuronal framework, in which a given lexical category representation can be used only once in the course of the analysis of a sentence. Chapter 12 addresses this issue and refines the notion of a neuronal grammar so that repeated elements can be processed. Furthermore, to provide evidence that neuronal grammars can compete with traditional grammars in the description of natural languages, it appears of particular interest to investigate whether the former are capable of processing string types that are outside the reach of finite-state grammars, but are of particular relevance in the description of natural languages, namely strings with center embeddings. This problem is taken as a test case for an extended version of neuronal grammar in Chapter 12.

The second reason is that it is always possible to introduce a rule such as (14) into an algorithm, such as the rule defined by (10)–(12).

(14) b (/∗/ e)

This implies that, in a chain *abcde* defined by (13), there would be an additional relationship between the second and the last elements (b and e). On the background of grammar (10)–(12), this would imply that an additional serial-order constraint is introduced. The number of strings the grammar accepts would become smaller.

There is no easy way to insert (14) into (13). Certainly, the new rule cannot be inserted in the same manner as (11) and (12) are inserted into (10). This is because all elements defined by (13) are already in the formula, and what would need to be added by (14) is an additional relationship between temporally and spatially distant elements. If the rule derivation is represented in a two-dimensional graph (or tree; see Chapter 7), the addition of (14) to the system leads to crossings of lines in the graph, thus violating a basic principle of many grammar theories. In neuronal grammar, the multidimensional relationships between elements of a syntactic string can be captured, although it is not easily captured by bracketed formulas or trees.

The emphasis is on that the circuit side has the same feature. Also, the circuit side of neuronal grammar is not subject to the restriction of two-dimensional graphs. The circuits are diagrams realizing all the entities specified in the formulas of a fragment of neuronal grammar. Each lexical element is represented as an input unit, each lexical category as a set of sequence sets, and each directed link defined by sequence formulas is realized by an asymmetrical reciprocal connection between sequence sets. In principle, the lines symbolizing neuronal links may cross, although one may believe that it is preferable to choose displays in which unnecessary crossings are avoided.

No general statements are made here about the generative capacity of neuronal grammars except the following: Rule (15) specifies the set of strings (16).

(15) b (a /∗/ c, d)
(16) {abcd, abdc}

The addition of a second rule, (17), increases the size of the set of possible strings quite considerably to the set listed under (18).

(17) a (/∗/ f)
(18) {afbcd, abfcd, abcfd, abcdf, afbdc, abfdc, abdfc, abdcf}

Instead of two string types eight would now be included. This is because only the slashes (not the commas) in valence formulas and rules indicate sequential order. Note that this is a major difference between dependency systems and neuronal grammars of the present formulation. In dependency systems, a rule of the form "b (a, ∗ , c, d)" implies sequential order of all four constituents so that only one string form – namely, abcd – would be generated or accepted.

To exclude some of the eight string types defined by Rules (15) and (17), additional rules would be necessary; for example, (19), which again restricts the set of possible string types.

(19) f (/∗/ c)
(20) {afbcd, abfcd, afbdc, abfdc}

The ability of a set of neuronal grammar rules to produce different sequences of the same lexical categories could be helpful for modeling properties of languages. There are sentence types in which the serial order of certain constituents is not constrained. Examples in English include double object constructions and sentences with transitive verbs that require particles, where the ordering of the two complements can be changed, as shown by example sentences (21)–(25).

(21) Betty puts a hat on.
(22) Betty puts on a hat.
(23) She rides on a bus to Bergamo.
(24) She rides to Bergamo on a bus.

In both cases, the sentences could be modeled by a rule such as (15) [repeated here as (25) for convenience].

(25) b (a /∗/ c, d)

Filling this abstract example with the concrete lexical categories used in (21)–(24) results in (26) and (27), respectively.

(26) V14p (N1 /∗ / N4, Vpart)
(27) V1PP (N1 /∗ / Prep1, Prep2)

In both cases, the verb requires two elements in its near future – accusative noun plus particle in one case and two different prepositions (Prep1, Prep2) in the other. The order in which these must follow is, however, not fixed in language use. To allow this same freedom in the formalism, valence formulas with free placing of the elements to the right of the asterisk can be advantageous.

As emphasized earlier, this formulation has some similarity to traditional syntactic theories, namely to dependency grammars and valence theory (Gaifman, 1965; Hays, 1964; Heringer, 1996; Tesnière, 1953). The underlying idea that members of lexical categories require complements (and can actually be defined as requiring complements) that are members of other lexical categories was expressed and systematically treated in Tesnière's work, but related ideas have been crucial in other grammar theories as well (e.g., categorical grammar; Ajukiewicz, 1936; Lambek, 1958, 1959). Formulas representing connections between input and sequence sets (assignment formulas) are analogous to assignment rules or lexicon rules defining the lexical categories of words and morphemes. Rules of this kind are probably specified in all major grammar theories. Descriptions of lexical category representations (valence formulas) are inspired by and are somewhat similar to statements specifying the valence of lexical categories and their dependencies, which are typical of dependency systems (see Chapter 7). There is no obvious analogon in grammar theories to the formulas representing connections between neuronal sets (sequence formulas).

11.4 Summary

In conclusion, the neurobiological grammar circuits proposed here have an equivalent formal description. This formalism is not equivalent to standard rewriting syntax or dependency theory, although it exhibits family resemblance with both. The main difference is that each formula refers to a neuronal device. Any attempt to simply reformulate the class of neuronal grammars in the languages of phrase structure or dependency grammars results in algorithms that produce trees with projection lines that cross or graphs with loops, something prohibited in these grammar theories. It may be that the present neurobiologically motivated proposal of syntactic algorithms can open new perspectives for grammar theories.

EXCURSUS TWO

Basic Bits of Neuronal Grammar

How does neuronal grammar operate? The following examples further illustrate the activation processes taking place in a network of neuronal sets during perception of congruent or grammatically well-formed and incongruent or ill-formed strings. This excursus aims to illustrate the principled difference in network dynamics between the processing of congruent and incongruent word strings, and further aims to introduce illustration schemes for network dynamics that are used in later sections of the book (see E3–E5 Chapters 11, 13).

Although the general mechanism of serial-order detection, mediated sequence detection by sequence sets, is simple, the interaction of several neuronal sets can become quite complex. To make their activity dynamics easy to overlook, two strategies are used to illustrate processes in grammar networks. One strategy is to list activity states of all sets contributing to the processing of a string at each point in time when a string element is present in the input and shortly thereafter. Activity dynamics are therefore presented in the form of tables. A second strategy is to present the simulations as animations. The animations, including illustrations of the three examples presented in this excursus, are available on the Internet at this book's accompanying web-page (http://www.cambridge.org).

E2.1 Examples, Algorithms, and Networks

Strings such as (1), (2), or (3) could be taken as examples for illustrating the function of a simple grammar network.

(1) Betty get up.
(2) Betty switches . . . on.
(3) The set ignites.

These strings are grammatical and one may envisage a grammar network to "accept" them. This means that the presentation of the word chains causes activity dynamics in the network that satisfies certain criteria. Criteria for string acceptance in neuronal grammar networks are discussed in Chapter 10 and are dubbed *satisfaction, visibility*, and *synchrony*. In this excursus, simple illustrations of network dynamics present in a neuronal grammar during and after presentation of strings of these types are presented and discussed.

The type of grammar network considered here accepts strings such as (1)–(3) because it includes two sequence detectors, one of which responds to the first and second elements of the respective string while the other responds to the second word followed by the third. It is clear that such sequence detectors would not only respond to these particular words, but also to sequences of elements of larger word groups (see Chapters 9–11). Although the first word pairs of sentences (1) and (2) could be detected by the same sequence set, different sequence sets would probably be involved in the processing of all other word-pair sequences.

If the order of these words is being changed, one may feel that the resulting strings are less common than the strings (1)–(3). This is illustrated by strings (4)–(9).

(4) Up get Betty.
(5) On . . . switches Betty.
(6) Ignites set the.
(7) Up Betty get.
(8) On . . . Betty switches.
(9) Ignites the set.

Not all of strings (4)–(9) are incorrect or grammatically deviant. However, it is assumed that there is one type of grammar machine that accepts strings such as (1)–(3) and does not respond in the same manner to (4)–(9). Again, this is for exposition purposes. It is possible to introduce additional network elements whose interaction may yield acceptance of other combinations of these words as well. For example, strings (5) and (8) may be accepted because of such to be added sequence detectors.

A network composed of neuronal sets that accepts the well-formed string (1) but rejects alternative combinations of the same words is shown in Figure E2.1. The grammar depicted in Figure E2.1 can be rewritten by the following set of formulas:

(10) Betty ↔ N (noun)
(11) get ↔ V (verb)

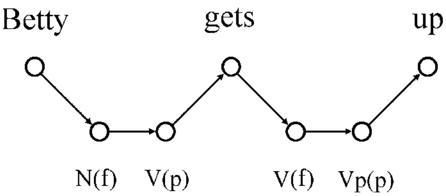

Figure E2.1. Elementary grammar network. Three word webs are connected by two sequence sets. All simulations in Excursus E2 are based on this circuit.

(12) up ↔ Vpart (verb particle)
(13) N (/∗/ f)
(14) V (p /∗/ f)
(15) Vp (p /∗/)
(16) N (f) ⟶ V (p)
(17) V (f) ⟶ Vpart (p)

There is correspondence between parts of the circuit and formulas (10)–(17). The relevant input units are defined with regard to their respective category labels by assignment formulas (10)–(12). Valence formulas (13)–(15) list the connection of each input unit to backward- and forward-looking sequencing units, and sequence formulas (16) and (17) have their network counterpart in the two horizontal arrows connecting two adjacent pairs of sequence sets.

E2.2 Grammar Circuits at Work

Table E2.1 presents a sketch of activity dynamics in the network during the processing of string (1), "Betty get up." Table E2.1 and all the following tables list the word webs at the top and are referred to by the word they represent. To the left and right of each word, the sequence sets connected to its word web are indicated. To the left of each word, its past-oriented sequence sets are listed, and future-oriented sets appear to the right.

The left row of the table lists the time steps of the simulation. The second column from the left gives the input to the grammar circuit – that is, the word stimuli activating network elements. Global activity states are indicated in the left column, also in capital letters.

For each time step (rows) and neuronal set (columns), the table matrix lists activity states. For each input unit, activity states are listed between slashes. For each sequence set connected directly to an input unit, activity states are listed either to the left (past sets) or the right (future sets) of the slashes. The activity states introduced in Chapter 10 are distinguished – that is, ignition (I), reverberation (R), priming (P), and rest (0). Full ignition

Table E2.1. Time course of activation and deactivation of input sets (or word webs) and sequence sets of Fig. E2.1 during processing of the string "Betty get up," which is congruent with the network structure. Lines represent time steps. Time steps and input to the network are listed on the left. In each column, activity states of an optional past-oriented sequence set (p; to the left of slashes), of a word web (between slashes), and of an optional future-oriented sequence set (f; to the right of slashes) are shown. Time steps between inputs are chosen so that activity spreads until stable states are reached. Letters indicate ignition (*I*), reverberation (*R*), priming (*P*), and rest (0). $I^\wedge$ indicates full unprimed ignition, and *P* and *R* denote the highest levels of priming and reverberation.

		/Betty/f	p/Get/f	p/Up/
0		/0/0	0/0/0	0/0/
1	Betty	/I^/0	0/0/0	0/0/
2		/R/P	0/0/0	0/0/
3		/R/P	P/0/0	0/0/
4		/R/P	P/P/0	0/0/
5	gets	/R/P	P/I/0	0/0/
6		/R/P	I/R/P	0/0/
7		/R/I	R/R/P	P/0/
8		/I/R	R/R/P	P/P/
9		/R/R	R/R/P	P/P/
10	up	/R/R	R/R/P	P/I/
11		/R/R	R/R/P	I/R/
12		/R/R	R/R/I	R/R/
13		/R/R	R/I/R	R/R/
14		/R/R	I/R/R	R/R/
15		/R/I	R/R/R	R/R/
16		/I/R	R/R/R	R/R/
17		/R/R	R/R/R	R/R/
18	DEACTIVATE	/0/0	0/0/0	0/0/

$(I^\wedge)$ of a word web is distinguished from primed ignition (I). Reverberation states are labeled by numbers indicating the rank of reverberation levels in the activity hierarchy, R denoting the highest activity level, R' denoting the next level below R, and R'' indicating the third highest level. Note that these labels differ from those used earlier (e.g., R_1, R_2), a change made to make the present tables easier to survey.

The time steps where input is given to the grammar network are chosen arbitrarily, and to avoid cross-talk of activity that depends on exact timing.

The sentences should be accepted regardless of whether they are spoken quickly or slowly, with or without pauses between words, and regardless of whether additional constituents are being inserted between any two of the words. Therefore, the crucial dynamics in the network should not depend on exact delays between word inputs. As a rule, several time steps in which activity is allowed to spread through the network lie between any two subsequent inputs. The circuits are always allowed enough time to settle into one of their stable states before a new input is computed.

Principles underlying activity dynamics have been explained in great detail in Chapter 10 and are mentioned here only as far as they become relevant in the individual derivations. The most important processes are as follows:

- Words in the input activate word webs (ignition I if primed; full ignition $I^\wedge$ if not).
- After its ignition, a set reverberates.
- Ignitions spread to adjacent sets that are already strongly primed.
- Reverberating word webs prime their future sets, which in turn prime their adjacent sequence set and word web.
- Full ignition of a set reduces the activity state of all other sets that are in the state of reverberation or priming.

E2.2.1 Simulation 1: Acceptance of a Congruent String

Table E2.1 presents activity dynamics caused in the grammar network by the congruent sentence (18).

(18) Betty get up.

This sentence is assumed to be consistent with the network structure in Figure E2.1. It can be considered to be grammatical in English if it is used to give a command or recommendation.

Input units are being activated one after the other at time steps 1, 5, and 10, respectively [following principle (A2) in Section 4.5]. At the same time, they ignite (I). After ignition of each input unit, its future-oriented sequence detector is being primed [cf. principles (A3ii), (A3iii)]. Two time steps later, priming is further transmitted to the right, where it affects input units of possible successor words (A3iv). This priming is important for later processing. Crucially, it makes possible the primed ignition I of the input unit activated by the next input word. When primed (rather than full) ignition takes place, the threshold regulation mechanism is not invoked. Thus, threshold regulation only occurs when the first input unit ignites (A2i, A5i). At time step 1, there are no preactivated sets at R_1 or P_1; therefore, threshold regulation has no effect.

After primed ignitions at time steps 5 and 10, waves of ignitions are initiated that spread to the left, where sequencing units are in the states of priming or reverberation (A3i, A4ii, A4iii). These backward activity waves are of utmost importance in the present network type because they establish coherence between-word webs directly connected by sequence sets. By this mechanism, the network establishes a syntactic match, the complement relation, between two words. Backward spreading of ignitions occurs between time steps 5 and 8, and, again, between time steps 10 and 16. The "backward transmission" of activity synchronizes input units. The last backward spreading wave establishes synchrony among all neuronal sets activated in the present simulation. It resets the reverberation levels of all participating units so that they can be assumed to reverberate in a coordinated and well-timed manner, or if spreading of the ignition wave is assumed to happen rather quickly, even almost synchronously.

Ignitions of the leftmost sequence unit in the diagram (representing the sentence initial proper name) at time steps 8 and 16 terminate the rightward wave of ignitions. This is so because there are no remaining preactivated units to which ignitions could spread and refractoriness applies (A1v). This deactivation indicates that all syntactic matching processes were successful.

Finally, note that the criteria for "string acceptance" are met. After time step 16, all input units are active together with their set of sequencing units. Input units are therefore *satisfied* (criterion 1). All sets reverberate at the highest level of R_1. Therefore, they are called *visible* (criterion 2). Finally, a coherent wave of ignitions has involved all of them. The input units are therefore *synchronized* (criterion 3). Because the criteria of satisfaction, visibility, and synchronization are reached, the string can be said to be *accepted* by the network, and the active representations are finally being reset to 0. This illustrates how the network accepts grammatical string (18), "Betty get up."

E2.2.2 Simulation 2: Processing of an Incongruent String

What would be the network response to an ungrammatical or less regular string? The word sequence (19) is not consistent with the network structure and processing of this string will therefore be used as an example of the processing of an "ungrammatical" word sequence.

(19) Up get Betty.

Intuitively, the network may detect the mismatch between string and network structure by assessing that words occurring in direct or indirect

Table E2.2. Time course of activation and deactivation of input sets (or word webs) and sequence sets of Fig. E2.1 during processing of the string "Up get Betty," which is incongruent with the network structure. For explanation, see Table E2.1. P' and R' indicate reduced activity levels, and P'' and R'' further decreased activity states.

		/Betty/f	p/Get/f	p/Up/
0		/0/0	0/0/0	0/0/
1	Up	/0/0	0/0/0	0/I^/
2		/0/0	0/0/0	0/R/
3		/0/0	0/0/0	0/R/
4	gets	/0/0	0/I^/0	0/R/
5		/0/0	0/R/P	0/R′/
6		/0/0	0/R/P	P/R′/
7		/0/0	0/R/P	P/R′/
8	Betty	/I^/0	0/R/P	P/R′/
9		/R/P	0/R′/P′	P′/R″/
10		/R/P	P/R′/P′	P′/R″/
11		/R/P	P/R′/P′	P′/R″/

succession in this string do not match. The physiological indicator of this fact in the simulation may be that presence of two string elements in the input fails to cause ignition and reverberation of the sequence set connecting them. If the string causes reverberation of all word webs, they should be left at different levels of reverberation and fail to elicit a final wave of backward ignitions synchronizing them. Table E2.2 illustrates how this is achieved by listing the detailed sequence of activations and deactivations caused by the input sequence.

The rejection of the string by the network – or failure to accept it – is based primarily on the principle that weak connections (from right to left in the network in Fig. E2.1) do not allow for priming (cf. principles A3iii and A3iv). Because the input units are not primed, each exhibits unprimed or full ignition $I^\wedge$ when its respective word occurs in the input, at time steps 1, 4, and 8 (principle A2). Because full ignition activates the threshold control mechanism, all activity states of other neuronal sets are lowered (A5i). This process is effective at time steps 5 and 9, where the highest activity levels P_1 and R_1 are lowered to P_2 and R_2, and these are decreased to P_3 and R_3. Finally, because backward spreading of activity requires that the neuronal sets to which activity spreads are at one of the highest levels of priming, no backward spreading is initiated.

In this case, not a single criterion from the three criteria required for string acceptance is met. First, because the "bridge" of primed sequence detectors is missing after the ignitions, no synchronization can arise. The activity levels finally achieved are also different from each other, because each element in the input elicited a full ignition, causing threshold regulation, which was in turn a result of the lack of priming. Consequent to the unprimed ignitions, threshold control is activated (A5i) so that the excitation levels of the representations of earlier string elements are changed to lower levels of priming. Therefore, two of the three input units finally reverberated at lowered levels of R and were thus not visible. Furthermore, the three input units ended up not satisfied – that is, they lack a full set of sequence detectors at the highest reverberation level at the end of the computation processes.

A situation not covered by the comments in Chapter 10 arises in the computations described in Table E2.2. At time steps 7 and 11, there are sets that should exhibit two different activity states at a time. For example, there is an input unit reverberating at R_2 (or R') that, in addition, is primed at a different level, P_1 (or P). Dealing with this situation is discussed in more detail in Chapter 12. In the tables in this Excursus, the level of reverberation is taken into account and the additional priming is ignored.

E2.2.3 Simulation 3: Processing of a Partly Congruent String

A slightly different behavior of the same network is shown in Table E2.3. Here, string (20) is processed.

(20) Up Betty get.

Again, this string is inconsistent with the network structure and should be taken as another example of an ungrammatical string. Although the network is not prepared to accept it as a whole, word string (20) includes a substring for which a sequence detector would be available. The ordered word pair "Betty get" is the initial part of congruent sentence (18). String (20) should therefore be partly acceptable for the circuit.

The misplaced initial element fails to cause priming of representations of subsequent elements. The ignition at time step 1 does not cause such priming of the neuronal set corresponding to the second word in the string. Therefore, a full ignition $I^\wedge$ occurs at time step 4. The reverberation level of the initially activated set is lowered as a result of the activation of threshold control.

Because the activation processes following time step 4 finally lead to the priming of the neuronal set representing the last element of the string in the input, its presence in the input causes a primed ignition at time step 8, followed by backward spreading of ignitions. This process can be considered

Table E2.3. Time course of activation and deactivation of input sets (or word webs) and sequence sets of Fig. E2.1 during processing of the string "Up Betty get," which is incongruent with the network structure. For explanation, see Table E2.1. *P′* and *R′* indicate reduced activity levels and *P″* and *R″* further decreased activity states.

		/Betty/f	p/Get/f	p/Up/
0		/0/0	0/0/0	0/0/
1	Up	/0/0	0/0/0	0/I^/
2		/0/0	0/0/0	0/R/
3		/0/0	0/0/0	0/R/
4	Betty	/I^/0	0/0/0	0/R/
5		/R/P	0/0/0	0/R′/
6		/R/P	P/0/0	0/R′/
7		/R/P	P/P/0	0/R′/
8	gets	/R/R	P/I/0	0/R′/
9		/R/R	I/R/P	0/R′/
10		/R/I	R/R/P	P/R′/
11		/I/R	R/R/P	P/R′/
12		/R/R	R/R/P	P/R′/

the network equivalent of a judgment that a substring of the stimulus is coherent.

The backward wave of ignitions cannot spread to the neuronal set of the first string element. Crucial for this is that there is, at time step 8, no priming of the sequence detectors that could serve as a bridge between the relevant input representations.

Final synchronization of all word webs therefore fails to take place. In this case, however, the synchronization and visibility conditions are met for the representations of the last two string segments. Nevertheless, two of the three input units are not being satisfied and there is no synchronization of representation of the well-formed string part and the misplaced particle. Therefore, the network does not accept the string. Comparison of simulations 2 and 3 shows that degrees of disagreement between network structure and serial order of strings can be reflected in the sequence of activation and deactivations in the grammar network.

These examples illustrate very elementary properties of a serial-order circuit. To demonstrate that neuronal grammar circuits can process relevant syntactic structures in a biologically realistic manner, it is necessary to consider more complex and therefore more demanding examples. This is done in Excursus E3.

EXCURSUS THREE

A Web Response to a Sentence

Simulationing sentence processing in grammar circuits is important because it shows the processes that postulated by a neuronal grammar to occur in the brain when congruent and incongruent word and morpheme strings are processed. The examples discussed in Chapter 10 and Excursus E2 were introduced to illustrate the working of the envisaged grammar machinery, the principles of which are less obvious from a more complex simulation. However, the earlier examples can be considered to be toy simulations because the strings under processing exhibit far less complexity than most sentences commonly used in everyday language.

It is therefore relevant to look at more complex examples of neuronal circuits that may be the basis of syntactic knowledge and syntactic processing. In this Excursus, a sentence discussed earlier in the context of conventional grammar models (see Chapter 7) is again the target. First, the algorithmic version of a neuronal grammar processing this and similar sentences is presented and the corresponding network described. Subsequently, activity dynamics caused in the grammar circuit by the sentence in the input are discussed. An animation of this simulation is available on the Internet at the books accompanying web page (http://www.cambridge.org).

We look first at sentence (1).

(1) Betty switches the machine on.

Putative syntactic structures possibly underlying the processing of sentence (1) are spelled out in the context of dependency and rewriting grammars (Section 7.5; Fig. 7.1). Also, a very tentative neurobiological circuit was proposed earlier that may underlie the processing of (1). This earlier first-order approximation at a grammar circuit must now be revised in the light of conclusions drawn in Chapters 10 and 11.

In contrast to the toy examples used in Excursus E2, sentence (1) exhibits nontrivial features, including the following:

- Discontinuous constituent made up by the verb and its particle
- Agreement between subject and verb suffix
- "Left branching" of the verb phrase.

As explained in Chapter 7, distributed words such as *switch … on* – called *particle verbs* here because the verb root requires a particle ("on") – are difficult to capture by syntactic tree structures. Transformations, feature transport through the branches of trees or other additional mechanisms must be assumed to make the relationship between the two distant lexical elements manifest in the syntactic representation. The same problem arises for the agreement between the proper name and the verb affix; again, two spatially or temporally separate elements (the proper name *Betty* and the verb suffix –(*e*)*s*) must be linked. *Left branching* refers to the fact that in a standard phrase structure representation of sentence (1) as depicted in Figure 7.1, the VP (or V″) node carries two downward branches, one of which (the left one) branches again. Multiple branches to the left have been proposed to give rise to computational problems. It is relevant to confront neuronal grammar networks with strings including discontinuous constituents, agreement, and left branching, because these features can pose problems to grammars and sentence processors.

Instead of postulating different syntactic mechanisms to capture relationships between adjacent and distant string elements, as most grammar theories do, neuronal grammar bases short- and long-distance dependencies on a single mechanism provided by sequence sets (cf. Chapters 9 and 10). Constructions necessitating leftward branches in the context of rewriting grammars may be analyzed differently within other grammar frameworks (cf. Fig. 7.1). The "flat" representations of neuronal grammar, in which all complement lexical items are each linked directly to the other, avoids the assumption of leftward branching, at least for many sentence types in which a conventional grammar may postulate such.

E3.1 The Grammar Algorithm and Network

As stressed earlier, depicting a grammar fragment as a network is not the only way of representing it. The network can also be rewritten as an equivalent set of formulas in which lines indicate connections between word webs and neuronal sets processing information in the past or future that make up lexical category representations. A neuronal grammar fragment that can process (1) is given next. Three sets of formulas – assignment, valence, and sequence formulas – make up the algorithmic version of the grammar fragment. For explanation of terms, see Chapter 7, and for an explanation of the formulas, see Chapter 11.

The following assignment formulas specify a set of connections between representations of morphemes and their respective lexical category representations composed of sequence sets. Lexical categories are labeled by abbreviations, and their full names are given in brackets.

(2) Betty ↔ Prop (nominative proper name)
(3) switch ↔ V (transitive particle verb)
(4) -es ↔ Vs (verb suffix)
(5) the ↔ Det (determiner)
(6) machine ↔ N (accusative noun)
(7) on ↔ Vpart (verb particle)

When extending the grammar to a larger vocabulary, it is necessary to clarify these categories, some of which can be defined rather widely. All proper names, for example, should be part of the proper name category. The category of proper names would be closely related to that of ordinary nominative nouns. The distinctive feature of the two would be the past-oriented sequence detector of the ordinary noun category that responds to a preceding determiner, which is usually present with nouns that are not proper names. Note that the distinction between proper and other names has syntactic implications, but can be defined on the basis of semantic criteria as well.

In contrast to very widely defined categories with many members, the verb suffix category has only a few members. These members would be the different realizations of the verb suffix signaling present tense, third person, and singular. In written English, this suffix can be *s* or *es*, and in spoken language, the *s* can be voiced or unvoiced and further context variants of the acoustic pattern of the suffix can also be distinguished. These variants of the morpheme could be realized in the model as overlapping sets, the abstract representation of the third-person singular present suffix being represented by the intersecting parts of these sets. (A similar proposal was made at the semantic level in Section 5.2.3 to model family resemblance.) Vpart would also be a very much restricted "lexical category." It would only include one verb particle because choice of a different particle would imply a different meaning and grammatical implication of the particle verb.

The grammar fragment includes a set of lexical category representations that can be defined as sets of sequence sets. The following valence formulas define the lexical categories in terms of the sequence sets that constitute them.

(8) Prop $(/{*}/\ f_1, f_2)$
(9) N $(p_1, p_2\ /{*}/)$
(10) V $(p\ /{*}/\ f_1, f_2, f_3)$
(11) Vs $(p_1, p_2\ /{*}/)$
(12) Det $(/{*}/\ f)$
(13) Vpart $(p\ /{*}/)$

Each letter *p* (past set) refers to a sequence set sensitive to words or morpheme from the lexical category preceded by words of a different category, and each *f* (future set) refers to a sequence detector sensitive to a word or morpheme of that category followed by an element of a different category. Indices are used if more than one past- or future-sensitive sequence set is present in any given lexical category set.

The reciprocal but asymmetric connections between all sequence sets are specified by the following set of sequence formulas:

(14) Prop $(f_1) \longrightarrow$ V (p)
(15) Prop $(f_2) \longrightarrow$ Vs (p_2)
(16) V $(f_1) \longrightarrow$ Vs (p_1)
(17) V $(f_2) \longrightarrow$ N (p_2)
(18) V $(f_3) \longrightarrow$ Vpart (p)
(19) Det $(f) \longrightarrow$ N (p_1)

Note that all past- and future-oriented sequence sets are connected. The sequence and valence formulas taken together include the information that lexical category representations are connected to each other directly.

The network equivalent to these formulas is presented in Figure E3.1. Each small circle represents a neuronal set. Word webs, the neuronal units processing words or morphemes, are listed in the top line with the word or affix they are specialized for indicated. Below each word web, linked directly by arrows, are the sequence detectors the word web is connected to. The set of sequence detectors connected directly to a word web represents the lexical category representation the word is part of. These direct connections between word web and sequence sets are specified by valence and assignment formulas. The horizontal connections between sequence sets correspond to sequence formulas (14)–(19).

E3.2 Sentence Processing in Syntactic Circuits

Table E3.1 displays activity dynamics of the network processing sentence (1). The reader is now invited to follow the activity waves through the network. The principles underlying the dynamics are the same as in earlier

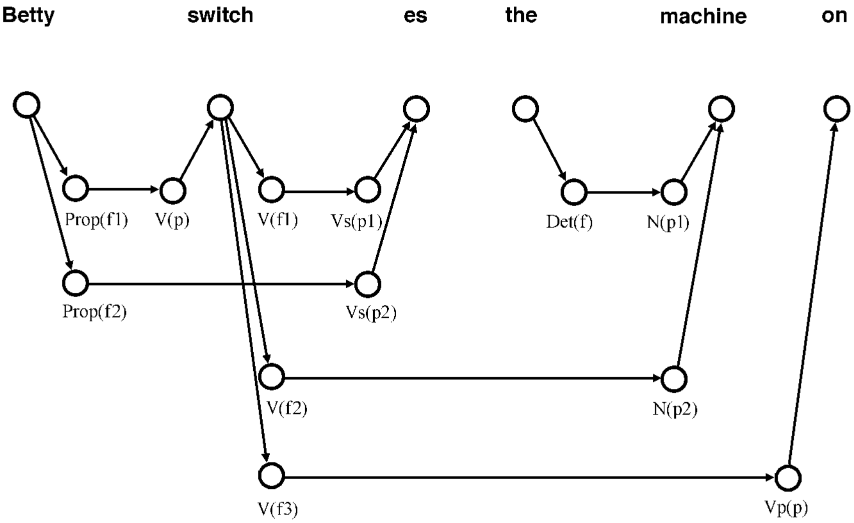

Figure E3.1. Fragment of neuronal grammar that processes and accepts the sentence "Betty switches the machine on." (An equivalent formalism is provided in the text.) Activation and deactivation dynamics of the network are shown in Table E3.1.

simulations (cf. Chapter 10 and Excursus E2). First, the global characteristics of the activation and deactivation processes in this simulation are briefly summarized. Finally, a few special features of this simulation are stressed.

E3.2.1 Global Characteristics

All activation and deactivation processes are described in a table organized in the same way as earlier activation tables of this kind. For a detailed description of how activation tables represent spreading activity through a network, see Section E2.2 in Excursus E2. The principles of neuronal grammar that yield the activation and deactivation processes have been discussed in great detail in Chapter 10 (e.g., in Section 10.7).

Each lexical element in the input activates its input unit. Here, the verb suffix – the third-person singular present suffix realized, in this case, as *-es* – is represented as a separate lexical element. Therefore, the five-word sentence activates six word or morpheme webs in a given order. The time steps at which word webs ignite are 1, 5, 10, 16, 20, and 29. As explained earlier, when a new item is presented, the time steps are chosen such that the network settles into a stable state before a new input is given. Each word or morpheme in the input causes an ignition of the word web representing it. Except for the first word, *Betty*, and the article *the*, the ignitions caused are primed ignitions, I, because the respective input units had already been preactivated. The other two ignitions are full ignitions, abbreviated as $I^\wedge$.

The ignitions change the activity states of other sets that are connected directly to them. The ignition itself and the reverberation state following each ignition produce priming of the directly connected sequence set sensitive to information about future events. Spreading of priming is indicated by changes of the inactive state, 0, to the state of priming, P. In Table E3.1, these changes are indicated to the right of word webs ignited by external input. Priming of strongly connected sets takes place after each ignition caused by input, except for the last one (time step 29). Priming can be effective up to the next word web connected to the reverberating input unit by sequence sets (*future sets*). Another way to express this would be to say that a word prepares the circuit to perceive a word from a category whose members frequently follow the word. However, if the next word web has more than one connected sequence set sensitive to information available in the past (labeled p plus index), all of its past sets need to be primed in order to cause priming of the word web (cf. Section 10.7). Ignitions also momentarily prime their directly connected sequence sets to the left. This temporary effect is not covered by Table E3.1, except for in the cases in which it has long-lasting consequences (traveling waves of ignitions).

An ignition caused by external input – the word or morpheme stimuli – leads to additional primed ignitions in directly connected sets that are already in a preactivated state (that is, sets that already reverberate or are in the state of priming at high levels). This leads to backward-traveling waves of activity, from right to left in the table, which are said to synchronize the sets involved. Backward activity spreading of primed ignitions occurs in the intervals of time steps 5–8, 10–14, 20–27, and 29–38.

There is also the possibility of indirect influence of an ignition on other neuronal sets via activation of threshold regulation processes. This becomes effective only after the determiner *the* has been presented, following time step 15. Activity levels of all preactivated sets in the network are reduced, from R_1 to the lower level R_2, or from P_1 to P_2, as a consequence. As in Excursus E2, the canonical activity levels R_1, R_2, R_3 . . . and P_1, P_2, P_3, . . . are given as R, R', R'', . . . and P, P', P'' in the table.

After the sentence had occurred in the input, activity reaches a stable state at stime step 39. At this point in time, the network has all its word webs or input units reverberating at the highest reverberation level R (also called R_1). The visibility condition is therefore satisfied. In addition, all of the word webs have highly active connected lexical category representations (i.e., sets of sequence sets. This means that the condition of satisfaction is also met. Finally, a terminal wave of ignitions spreading throughout all activated sets

Table E3.1. Time course of activation and deactivation of word webs and sequence sets of Figure E3.1 during processing of the string “Betty switches the machine on,” which is congruent with the network structure. Lines represent time steps. Time steps and input to the network are listed on the left. In each column, activity states of one or more past-oriented sequence sets (p; to the left of slashes), of a word web (between slashes), and of future-oriented sequence sets (f; to the right of slashes) are shown. Time steps between inputs are chosen so that activity spreads until stable states are reached. Content words are abbreviated (B. – Betty, s. – switch, m. – machine). Letters indicate ignition (*I*), reverberation (*R*), priming (*P*), and rest (0). *I*^ indicates full unprimed,ignition. *P* and *R* denote the highest levels of priming and reverberation and *P*′ and *R*′ indicate lower activity levels.

	/B./f_1,f_2	p/s./f_1,f_2,f_3	p_1,p_2/-s/	/the/f	p_1,p_2/m./	p/on
1 Betty	/I^/0,0	0/0/0,0,0	0,0/0/	/0/0	0,0/0/	0/0/
2	/R/P,P	0/0/0,0,0	0,0/0/	/0/0	0,0/0/	0/0/
3	/R/P,P	P/0/0,0,0	P,0/0/	/0/0	0,0/0/	0/0/
4	/R/P,P	P/P/0,0,0	P,0/0/	/0/0	0,0/0/	0/0/
5 switch	/R/P,P	P/I/0,0,0	P,0/0/	/0/0	0,0/0/	0/0/
6	/R/P,P	I/R/P,P,P	P,P/0/	/0/0	0,0/0/	0/0/
7	/R/I,P	R/R/P,P,P	P,P/P/	/0/0	P,0/0/	P/0/
8	/I/R,P	R/R/P,P,P	P,P/P/	/0/0	P,0/0/	P/P/
9	/R/R,P	R/R/P,P,P	P,P/P/	/0/0	P,0/0/	P/P/
10 es	/R/R,P	R/R/P,P,P	P,P/I/	/0/0	P,0/0/	P/P/
11	/R/R,P	R/R/P,P,P	I,I/R/	/0/0	P,0/0/	P/P/
12	/R/R,I	R/R/I,P,P	R,R/R/	/0/0	P,0/0/	P/P/
13	/I/R,R	R/I/R,P,P	R,R/R/	/0/0	P,0/0/	P/P/
14	/R/I,R	I/R/R,P,P	R,R/R/	/0/0	P,0/0/	P/P/
15	/R/R,R	R/R/R,P,P	R,R/R/	/0/0	P,0/0/	P/P/
16 the	/R’/R’,R’	R’/R’/R’,P’,P’	R’,R’/R’/	/I^/0	P’,0/0/	P’/P’/
17	/R’/R’,R’	R’/R’/R’,P’,P’	R’,R’/R’/	/R/P	P’,0/0/	P’/P’/
18	/R’/R’,R’	R’/R’/R’,P’,P’	R’,R’/R’/	/R/P	P’,P/0/	P’/P’/
19	/R’/R’,R’	R’/R’/R’,P’,P’	R’,R’/R’/	/R/P	P’,P/P/	P’/P’/
20 machine	/R’/R’,R’	R’/R’/R’,P’,P’	R’,R’/R’/	/R/P	P’,P/I/	P’/P’/
21	/R’/R’,R’	R’/R’/R’,P’,P’	R’,R’/R’/	/R/P	I,I/R/	P’/P’/
22	/R’/R’,R’	R’/R’/R’,I,P’	R’,R’/R’/	/R/I	R,R/R/	P’/P’/
23	/R’/R’,R’	R’/I/R’,R,P’	R’,R’/R’/	/I/R	R,R/R/	P’/P’/
24	/R’/R’,R’	I/R/I,R,P	R’,R’/R’/	/R/R	R,R/R/	P’/P’/
25	/R’/I,R’	R/R/R,R,P	R’,I/R’/	/R/R	R,R/R/	P/P’/
26	/I/R,R’	R/R/R,R,P	R’,R/I/	/R/R	R,R/R/	P/P/
27	/R/R,I	R/R/R,R,P	I,R/R/	/R/R	R,R/R/	P/P/
28	/R/R,R	R/R/R,R,P	R,R/R/	/R/R	R,R/R/	P/P/
29 on	/R/R,R	R/R/R,R,P	R,R/R/	/R/R	R,R/R/	P/I/
30	/R/R,R	R/R/R,R,P	R,R/R/	/R/R	R,R/R/	I/R/
31	/R/R,R	R/R/R,R,I	R,R/R/	/R/R	R,R/R/	R/R/
32	/R/R,R	R/I/R,R,R	R,R/R/	/R/R	R,R/R/	R/R/
33	/R/R,R	I/R/I,I,R	R,R/R/	/R/R	R,R/R/	R/R/
34	/R/I,R	R/R/R,R,R	R,I/R/	/R/R	I,R/R/	R/R/
35	/I/R,R	R/R/R,R,R	R,R/I/	/R/R	R,R/I/	R/R/
36	/R/R,I	R/R/R,R,R	I,R/R/	/R/R	R,I/R/	R/R/
37	/R/R,R	R/R/R,R,R	R,R/R/	/R/I	R,R/R/	R/R/
38	/R/R,R	R/R/R,R,R	R,R/R/	/I/R	R,R/R/	R/R/
39	/R/R,R	R/R/R,R,R	R,R/R/	/R/R	R,R/R/	R/R/
40 DEACTIVATE	/0/0,0	0/0/0,0,0	0,0/0/	/0/0	0,0/0/	0/0/

(time steps 29–38) follows the last stimulation by an element in the input. The synchronization condition is therefore met as well. The sentence is in accord with the structure of the circuit. Therefore, the activations induced by the sentence are being terminated by the postulated deactivation process.

E3.2.2 Specific Features

It may be worthwhile to consider the following special features of the processing of sentence (1) that address the three specific features of its syntactic structure – namely, the included discontinuous constituent, the agreement exhibited by two of its constituents, and the left branching attributed to it by a phrase structure grammar (PSG).

The discontinuous constituent *switch... on* does not pose problems to this processing device because the particle is bound to the verb by the same kind of connection assumed between the verb and its other complements. This is related to the fact that in a neuronal grammar, there is no motivation to prohibit what would be crossings of projection lines in a PSG. Crossings of projection lines are prohibited in rewriting systems, although there is probably no biological motivation for this. Following time step 5 of this simulation, the verb stem primes its connected sequence detectors that also link it to a verb suffix potentially activated in the future. Priming of the verb suffix representation becomes effective at time step 8, so that its activation at time step 29 is primed ignition.

Also, the fact that choice of the verb suffix depends on both the subject and the verb is transformed naturally into the network. Connections are postulated from both the proper name Prop to the verb suffix Vs, and, in the very same way, from the verb V to the verb suffix Vs. Neuronal grammar treats agreement between verb and verb suffix in the same way as any other dependency (cf. Chapter 9). Again, the point here is that dependency relations need not be two dimensional, but can have as many dimensions as required to describe cooccurrence of morphemes in sentences. In this simulation, the process most important in this context – the priming of the verb suffix representation by two sequence sets, one of them sensitive to the earlier proper name and the other to the preceding verb – becomes effective at time step 7.

The determiner *the* preceding the noun *machine* does not have a primed neuronal representation when this word occurs in the input. This is because the article's web linked only to the noun's web, and the noun web has not yet been activated. A full, rather than primed, ignition is therefore achieved by the determiner that reduces activity states of already active representations. There is a danger that the already active and the newly activated representations will fail to be linked by synchronization processes, a situation that sometimes arises when a neuronal grammar processes a sentence structure exhibiting left branching in a phrase structure representation.

Processing part of the left-branching structure of the string initially induces full ignition because the representation of the article lacks any past set and, therefore, there is no possibility to prime it before it ignites. At time step 15, the initial sentence fragment *Betty switches* has been processed as a coherent unit, that is, its representation has been synchronized and is visible. As a result of the unprimed ignition $I^\wedge$ caused by the article in the input at time step 16, threshold regulation is activated and the visible elements are pushed to a reduced level of reverberation R' (or R_2). The activity states of most sets stay at a reduced level until the structure *the machine* is completed. At time step 20 and later, however, synchrony is established between the word sets of the noun and article that reverberated at different levels of R (R). The ignition wave initiated by *machine* in the input finally spreads to the representations of both the initial sentence parts and the left-branching parts, thus synchronizing all the then-active sets. This process takes place during time steps 20–28 (cf. Table E3.1).

E3.3 Discussion of the Implementation and Simulation

This proposal of neuronal grammar circuits is but one step toward a brain-compatible serial-order mechanism that could be relevant for syntactic processing in the human brain. It is relevant to discuss future alternatives. The following issues may require additional labor.

Although the network presented in Figure E2.1 can process string (1) and may equally process versions of this string in which words and morphemes have been replaced by other members of the respective lexical categories, the mechanism must be further elaborated and differentiated in many ways, as it contains some shortcomings, and different perspectives can be offered.

Clearly, the formulation of the present grammar allows for different word order, not just for the one in sentence (1). Although the formulation can also model sentence (20), it does not "accept" strings in which complements are missing such as (21), or strings with different word order such as (22).

(20) Betty switches on the machine.
(21) Betty switches on machine.
(22) On switches Betty the machine.
(23) Betty switch the machine on -es.

However, the outlined network would also accept strings one would not want accepted. Such "false positives" would include (23), for example. To exclude their acceptance, the network must be refined.

One perspective for further development is offered by the introduction of connections of different strengths. Only one type of strong (forward) and weak (backward) connection has been set up in the grammar networks discussed in detail here. The present proposal could be further developed to capture the fact that the verb suffix must occur directly after the verb stem, whereas accusative noun phrase and particle can be placed later in the string. In principle, differential connection strength could realize this. Connection strength 1.0 could be postulated for the link of V to Vs, but a weaker link, 0.8, could be assumed between V and Vpart. The stronger links could then be effective earlier and the weaker ones only later. This would not remove the advantage of allowing for free placement of complements [cf. sentences (1) and (20)] because connection strengths between a constituent and two of its complements can always have the same value, thus leaving open which constituent must occur first. Because differential connection strengths would make the present computations more complex, this possibility is not dealt with here in more detail.

One may argue that it might be a caveat of the approach proposed here that not only the standard word order is possible in English, and this problem is even more virulent for other languages. The simple solution suggested, however, is that different grammatical word orders of the same or similar sets of constituents can be processed by different sets of sequence feature detectors. An active and a passive sentence and a main clause and a subordinate clause would therefore be characterized by distinct sets of sequence detectors. As argued in Section 10.6.4, it is not costly to add sequence sets for processing strings with different word order.

One may argue that the philosophy underlying this kind of sentence processing has some relationship to standard grammatical theories and that there is nothing in these computations that a modern grammar could not do. This is clearly correct. By introducing concepts such as slash categories, or operations such as movement or tree adjoining, the capacity of a grammar algorithms have been greatly augmented, so that the claim cannot be made that neuronal mechanisms cover strings that conventional grammars fail to process.

A major difference between conventional syntax theories and neuronal grammar is, of course, that the present proposal attempts to specify brain mechanism on a concrete level. A further difference to a many standard grammatical analyses lies in the fact that these grammars compute a tree representation as an output, whereas the neuronal automaton has a group of satisfied, visible, and synchronized neuronal sets activated and finally deactivated. It is possible that this is, at a certain level of abstraction, equivalent to one or more nonneuronal grammar algorithms. Chapter 11 claims that a certain modification of dependency syntax is equivalent to grammar circuits. It is not impossible that slight modification of other grammar types lead to descriptions that are compatible with the present concept of a neuronal grammar.

A further criticism of neuronal grammar could be as follows: It could be argued that the processes in grammar networks lead only to the activation and grouping of representations of lexical items. This could be considered to be similar to listing the lexical items included in a sentence together with one bit of information about whether the string was grammatical. It may be claimed that too much information about sentence structure and hierarchical relationships between lexical elements would be lost if computations were preformed in this manner. The argument may therefore be that a tree representation is preferrable because it preserves more structural properties of the morpheme string – in particular, its hierarchy.

However, this putative critique does not appear to be appropriate. In grammar circuits, word representations become active together with one of their respective lexical category representations and their neuronal links. Together with the activity dynamics of neuronal sets, these provide a structural characterization of the perceived string in much the same way as a tree graph, but with the important difference of not being restricted to two-dimensional representations. Furthermore, additional information becomes available in the course of sentence processing. During the computations in the circuits, information about relationships between lexical items is revealed by the process of synchronization and the level of reverberation of the sets. From the computation summarized in Table E3.1, it is clear that the first three elements of the sentence belong together from the beginning because they exhibit synchronization after the verb suffix leads to ignition of its input unit. Also, the close relationship between noun and article becomes visible in the adjacent ignitions of their input and sequencing units. In the final synchronization process initiated by the last word in the input, information about dependency relations between lexical items is revealed by the order in which their representations ignite. Thus, it appears that the present framework is quite capable of revealing information about dependency relations in sentence structures, although the tree structure hierarchies are usually not assumed in this type of grammar approach.

An additional advantage of this approach is that it offers a clear-cut description of the sequence of activation and deactivation processes, the sequence of groupings of constituents, and the temporal order in which synchrony is being achieved in the course of the processing of a sentence. Structural information about a sentence therefore unfolds over time in the grammar circuit.

CHAPTER TWELVE

Refining Neuronal Grammar

This chapter takes the concept of a neuronal grammar developed earlier as a starting point. The earlier proposal is partly revised and extended to cope with problems put forth by natural language phenomena.

The three phenomena considered here are as follows:

- The distinction between a constituent's obligatory complements and its optional adjuncts.
- The multiple occurrence of the same word form in a string.
- The embedding of sentences into other sentences.

All three issues have been addressed occasionally before; however, they were not treated systematically in the context of neuronal grammar. This requires its own discussion because the necessary extensions lead to a substantial revision of the previous concept of neuronal grammar.

The gist of the revision briefly in advance is as follows: In earlier chapters, the relationship between sequence detectors and words in the input was assumed to be static. The word web ignites and then, successively, one set of its sequence detectors is recruited according to relationships the word exhibits to words in its context, as they are manifest in regular co-occurrence of lexical category members. Each word would recruit one of its connected lexical category representations. However, if each word form were attributed to one lexical category, it would be impossible to model the situation in which one word occurs twice in different grammatical roles in a sentence. In this case, the single word representation must store the knowledge about the two occurrences, and the knowledge about the syntactic role the particular word occurrences played in their respective contexts as well. Section 12.2 discusses a preliminary solution to this dilemma. The idea here is that each word representation can exhibit several different activation states at a time. Two different waves of activity spreading through the same neuronal set may allow for storing the fact that a given item was present twice. The proposed mechanism can even operate if a lexically ambiguous word is used twice in a sentence, as member of two different lexical categories.

An even more radical use of the neuronal grammar machinery is offered in Section 12.3, where, finally, center embeddings are addressed. The solution explored here is that links between word webs and sequence sets are fully dynamic. The envisaged neuronal basis of the dynamic link is the traveling of a coherent wave of activity that reverberates back and forth between neuronal sets.

At the start, an important linguistic distinction between complements and adjuncts is reviewed and a few thoughts about its neuronal basis are added (Section 12.1).

12.1 Complements and Adjuncts

In Chapter 10, lexical elements were considered as units that "require" others. If one lexical element is present in a string, its complements can also be expected, with the converse also true – the "dependent" item also requires the "governor." This view gave rise to the concept of *mutual dependence* that avoids the hierarchy implication. Nominative noun and verb are mutually dependent. This is implemented by a reciprocal connection, with stronger links in the direction of the normal succession of the elements (from nominative noun to verb in English) than in the reverse direction. The postulate of reciprocal connections between neuronal sets α and β is based on the following assumptions:

- Large cortical neuron populations are connected to each other reciprocally.
- Strong correlation characterizes the occurrence of the words or morphemes represented by sets α and β.

These issues are addressed in earlier chapters (e.g., Chapter 2, in which the reciprocity assumption is discussed; Chapter 10, in which mutual dependence is based on reciprocal asymmetric connections).

In contrast to the case in which two types of lexical elements are correlated and usually cooccur, there are cases in which the occurrence of one lexical element depends on the occurrence of the other, but the reverse is not true. Whereas a noun (or proper name) and a verb usually co-occur, as in (1), the relation between a relative pronoun and a noun/proper name should probably be characterized as follows: If one occurs, the other is usually also present, but the converse is not necessarily so.

(1) Hans comes....
(2) Hans who....

The relative pronoun *who* can be added freely to any proper name referring to a human being, but the relative pronoun necessitates the noun or a similar element to which it can adjoin. If there is no noun or related item, no relative pronoun can occur. The nonobligatory, to be a freely added element is called a *free adjunct*. The distinction between obligatory complements and free adjuncts is common in grammar theories (see Ajukiewicz, 1936; Haegeman, 1991; Hays, 1964; Tesnière, 1953).

Note again the asymmetry between the adjunct and the noun: The word *who* used as a relative pronoun requires the occurrence of a noun or something similar (e.g., proper name, pronoun). The probability of finding *who* after a noun is generally low, but if a relative pronoun *who* is present, it is almost certain that also a noun or noun-like element (e.g., personal pronoun, proper name) is in the string.

How would this characteristic asymmetry of adjuncts be realized in a neuron circuit? Would it be realized through the same type of reciprocal and directed connection as postulated for the links between mutual complements?

It is well established that coactivation of a pre- and postsynaptic neuron leads to synaptic strengthening (see Chapter 2). However, presynaptic activation and postsynaptic silence can reduce synaptic weights dramatically. This process is called *homosynaptic long-term depression* (see Tsumoto, 1992). The reverse situation can have a similar effect as well. In this case, presynaptic silence while the postsynaptic neuron fires also reduces weights, a process called *heterosynaptic long-term depression*. Both processes, homosynaptic and heterosynaptic long-term depression, can lead to a reduction of synaptic weights between two connected neurons. It is still under discussion which of the two processes produces stronger synaptic weight reduction. However, there are theoretical arguments why, in a network made up of distributed and strongly connected sets of neurons, one of the processes should have a stronger effect than the other. From computer simulations, it became apparent that for achieving stable functioning of a system of cell assemblies, it was necessary to introduce a "post- not pre-" heterosynaptic rule of synaptic weight reduction (Hetherington & Shapiro, 1993). One may therefore propose that presynaptic silence with postsynaptic firing leads to the most pronounced reduction of synaptic weights. The assumption that long-term depression is more effective if it is heterosynaptic (post- not pre-) gives rise to a prediction on the neuronal implementation of free adjuncts.

If a neuronal element β is always active when α is active, but there are many cases in which α is active while β is silent, then the connections from α to β are assumed to suffer more from synaptic weight reduction than those backward from β to α. This postulate is based on a "pre- not post-" rule of weakening of synaptic connections. We assume here that the situation of a lexical element of type b relying on the presence of another one of type a – as the relative pronoun relies on the noun, or the *who* relies on *Hans* in example (2) – is modeled in the network by unidirectional connections from neuronal sets β to α. The idea here is that, although reciprocal connections between sets can be assumed to be a default (see discussion in Chapter 2), learning that a can appear without b, whereas b most frequently occurs together with a, causes one of the two links to become ineffective.

The unidirectional links imply that ignitions can spread only from the adjunct representation to the "mother", but no significant amount of activity can be transmitted in the opposite direction, from the mother representation to the adjunct. The adjunct–complement distinction is an important one for language functioning, and its putative neurobiological realization outlined here plays a role in the syntactic circuits elaborated in this chapter and the subsequent excursuses.

12.2 Multiple Activity States of a Neuronal Set

Multiple occurrence of lexical items and other syntactic objects is a genuine characteristic of natural language. To allow words to occur repeatedly in a string is at the heart of the view on language proposed by most mainstream linguistic theories because, from this perspective, each language is considered to be a system that is enumerable recursively. This implies that rules can be applied repeatedly in the analysis or synthesis of a sentence, and it also implies that, in principle, the number of possible sentences is unlimited. One may want to call this the *linguistic infinity position* or *linguistic infinity view*.

The linguistic infinity view implies that sentences are, in principle, allowed to have infinite length. Also, because the base vocabulary of each language is considered to be finite, this view further implies that lexical items are allowed to occur repeatedly within sentences, with no upper limit, in this case, for the number of repetitions. Therefore, from this linguistic infinity perspective, it appears imperative to specify possible mechanisms underlying processing of multiple occurrences of morphemes. From a more realistic viewpoint, it also appears advantageous to have a mechanism at hand that allows one to have multiple occurrences of the same lexical element in well-formed strings.

12.2.1 The Concept of Multiple Reverberation

How could a neuronal grammar process a sentence in which the word form *switch* occurred twice? One would probably assume that the neuronal representation of the word would become active for a second time when the second occurrence of the word occurs in the input. However, to allow for processing of sentence-related information at a later point in time, information about both occurrences must be stored, and it would also be necessary to store not only that they occurred but also which syntactic function they had in their respective positions. The nontrivial task for a neuronal grammar is illustrated using sentence (3).

(3) Betty switches the switch on.

In the context of this sentence, the word form *switch* can be assumed to be used as a member of two lexical categories specified by the Rules (4) and (5).

(4) V14p (Np /*/ Vs, N4, Vp)
(5) N4 (V, Ar /*/)

Thus, the word form is used both as transitive particle verb and as accusative noun. Now the problem to be solved is at least two-fold: The network must store the following facts:

(i) The lexical item occurred twice.
(ii) The lexical item occurred as member of two specific lexical categories.

In the networks defined here, a symbol in the input activates its input unit and a set of sequence detectors. If a word form occurs twice but in different syntactic functions – that is, as member of different lexical categories, a larger number of sequence detectors are activated in comparison to the case of the same element occurring only once. However, according to principle (A6) in Section 10.7, simultaneous ignition or reverberation at R_1 of two different lexical category representations connected to the same input unit is not possible. Therefore, it is unclear how the network would store items (i) and (ii). To allow processing of repeated use of the same word form, a fundamental revision of the neuronal model appears to be necessary.

One possible solution is that individual neuronal sets support multiple simultaneous reverberations. The proposal is that several distinct waves of well-timed activity can run simultaneously through a given set.

After its ignition, a neuronal set is proposed to reverberate at R_1. One way to model reverberation is by using the paradigm of a huge neuron loop in which not individual neurons but groups of neurons are connected in such a way that the groups form a circle, or a more complex circuit including many loops (Chapter 8). If there are directed connections between the neuron groups, an active group activates the next group that, in turn, activates the next, and so on. Circulation and reverberation of activity results in such a model of *reverberatory synfire chains*. Important assumptions are that reverberations last for tens of seconds, and that two or more separate reverberatory waves of activity in the same network do not interfere with each other.

A few considerations on putative mechanisms underlying multiple reverberation may appear relevant here. A neuronal set has been proposed to have an ordered inner structure, in the manner of a reverberatory synfire chain (Chapters 8 and 10). A reverberating wave of activity in a neuronal set may emerge after full activation of all neurons of the set because after the ignition, refractory processes in strongly active neurons reduce the level of activity, ultimately leaving only a small subset of neurons active. If an igniting assembly includes a very strongly connected subset of neurons, its *kernel* (Braitenberg, 1978a), these neurons are the first to become strongly active, and therefore fatigue (e.g., refractoriness) is first present in this subset. Shortly thereafter, the kernel neurons are the first to become active again, because their refractoriness is already over at the time when most other set neurons are still refractory. Therefore, reverberations should start in the kernel and spread from there through the rest of the set. This would be the situation in a neuronal set activated only once.

The putative mechanism of multiple reverberation is the following. If a wave of activity already reverberates in the set, a second ignition as a result of external stimulation does not disturb the first wave, but causes a second distinct wave that also reverberates in the set at the same frequency, but at a different phase. If, at the time of the second ignition, the set is in the reverberating state, one group of neurons must be active and another set – the subpart of the synfire chain that was active just one time step before – must be in its refractory period. Therefore, it cannot take part in the second ignition. For this reason, there is a group of neurons in the set that does not participate in the second ignition. These neurons in the refractory group are among the first that become active again after ignition ceases. If an ignition usually results in one wave of reverberating activity starting in the kernel, in this case there are two such waves that emerge after the second ignition of an already reverberating set. Although wave one would be initiated in the kernel, the second wave would, so to speak, result from the "footprint" of the earlier reverberation.

It may be fruitful to explore these putative mechanisms further using computer simulations of reverberatory synfire chains (Bienenstock, 1996). Such simulation can further explore the idea that, if a neuronal set is conceptualized as a reverberatory synfire chain in which well-timed waves of activity resonate (see Chapter 8), more than one wave can reverberate at a given point in time within any given set. The application of this idea to grammar circuits is explored further in Excursuses E4 and E5, in which more simulations of the processing of sentences is discussed. In the following considerations, it is assumed that input units – word and morpheme webs – can exhibit multiple states of reverberation as a result of multiple stimulation by external input. Multiple activity states can also be postulated for sequence sets.

12.2.2 Some Revised Principles of Neuronal Grammar

A *neuronal set* can now be redefined as an entity that can exhibit several simultaneous states of reverberation, priming, and ignition at any point in time. An ignition induced by external input causes a wave of reverberation within the input unit and the number of ignitions during a given period determines the number of its reverberations. It is emphasized that there is, at present, no evidence available supporting this postulate. This postulate – or an alternative basis of an account of the processing of multiple occurrences of the same language element – is necessary in any approach to grammar mechanisms in the brain.

As a consequence, the ignition of an already reverberating input unit should not result in a primed ignition but in a new full ignition instead. This makes revision of the principles outlined in Section 10.7 unavoidable. Furthermore, it is not clear how the idea of multiple reverberations of input units matches that of threshold regulation as specified by axiom (A5) in Section 10.7. If full (unprimed) ignition $I^\wedge$ reduces R_i to R_{i+1}, an additional $I^\wedge$ further diminishes all already established reverberation to R_{i+2}. Whereas an input unit that is already in the state of reverberation usually exhibits unprimed full ignition, it still exhibits primed ignition if being in a state of priming at a high level.

To exclude overly strong activity because of multiple reverberation, a set is only allowed one wave of reverberation at the highest level R_1 at any given time. If an ignition takes place while the set reverberates at R_1, two waves of activity emerge at R_1 and R_2, respectively. Next, input units are assumed to produce multiple reverberation as a consequence of ignition by external input if they have already been in a state of reverberation R_i before the ignition. Axiom (A2) must now be modified and reformulated as (A2′):

(A2′) Activity change in an input unit S caused by external stimulation

(i) $S(0,t),\ S(E,t) \Rightarrow S(I^\wedge,t+1)$
(ii) $S(P_1,t),\ S(E,t) \Rightarrow S(I,t+1)$
(iii) $S(P_i,t),\ S(E,t) \Rightarrow S(I^\wedge,t+1)$
if $i > 1$
(iv) $S(R_1,t),\ S(E,t) \Rightarrow S(I\ R_2,t+1)$
(v) $S(R_i,t),\ S(E,t) \Rightarrow S(I^\wedge\ R_i,t+1)$
if $i > 1$

where $S(x_1, x_2, \ldots x_h, y)$ is a predicate with $h+1$ places indicating the h activity states ($h > 0$) of neuronal set S at time y. x_1 can take the values 0, I, or any level of R or P, and x_2 to x_1 can only take levels R_i or P_i, $i > 1$. Thus, multiple activity states can be expressed, for example, by (6).

(6) $S(I\ R_3\ R_7,\ t)$

This means that at time t, set S simultaneously ignites and reverberates at R_3 and R_7.

There is, in principle, no upper limit for the number h of simultaneous activity states. If axioms (A2′iv) or (A2′v) are allowed to be applied h times to the same input unit, the unit exhibits h activity states simultaneously. Although an abstract version of the model may postulate high numbers of h, a realistic model would probably limit the number of each word web's simultaneous activity states to a few.

The model implies that any set receiving more than one inputs through a strong connection from other sets can exhibit multiple states of priming. This is so for the following reasons:

(i) An input unit S_p can reverberate at different levels.
(ii) If S_q receives strong connections from set S_p, priming of S_q is, as a consequence of different levels of reverberation, stronger or weaker as well.

This is implied by the earlier formulation of neuronal grammar in Chapter 10 (see also Excursus E4). According to principles (A3iii) and (A3iv), reverberation and priming of a set leads to priming of a strongly connected set, and, if reverberation is at levels R_1, R_2, and so on, the priming levels are, accordingly, adjusted to P_1, P_2, and so on. Thus, levels of activity are, so to speak, transported through strong connections. Briefly, the following modifications of principles (A3) and (A4) are proposed:

(A3′) Activity changes caused in S_q through a connection between S_p and S_q

(i) $S_p\ (I,t),\ S_q\ (0,t) \Rightarrow S_q\ (P_1,t+1)$ only if $S_p \longrightarrow S_q$

(ii) $S_p\ (I,t),\ S_q\ (R_i,t) \Rightarrow S_q\ (I,t+1)$ only if $i < 3$

(iii) $S_p\ (I,t),\ S_q\ (R_i,t) \Rightarrow S_q\ (P_1\ R_i,t+1)$ only if $S_p \longrightarrow S_q$ and $i > 2$

(iv) $S_p\ (I,t),\ S_q\ (R_i\ R_j,t) \Rightarrow S_q\ (I\ R_j,t+1)$ only if $S_p \longrightarrow S_q$, $i < 3$, and $j > 2$

(v) $S_p\ (R_i,t),\ S_q\ (0,t) \Rightarrow S_q\ (P_i,t+1)$ only if $S_p \longrightarrow S_q$

(vi) $S_p\ (R_i,t),\ S_q\ (R_j,t) \Rightarrow S_q\ (P_i\ R_j,t+1)$ only if $S_p \longrightarrow S_q$, and $i \neq j$

(vii) $S_p\ (R_i,t),\ S_q\ (R_i,t) \Rightarrow S_q\ (R_i,t+1)$ only if $S_p \longrightarrow S_q$

(viii) $S_p\ (P_i,t),\ S_q\ (0,t) \Rightarrow S_q\ (P_i,t+1)$ only if $S_p \longrightarrow S_q$, and S_p is a sequence set

(ix) $S_p\ (P_i,t),\ S_q\ (P_j,t) \Rightarrow S_q\ (P_i\ P_j,t+1)$ only if $S_p \longrightarrow S_q$, S_p is a sequence set, and $i \neq j$

(x) $S_p\ (P_i,t),\ S_q\ (P_i,t) \Rightarrow S_q\ (P_i,t+1)$ only if $S_p \longrightarrow S_q$, S_p is a sequence set

(A4′) Activity changes caused in S_r through connections with S_p and S_q

(i) $S_p\ (I,t),\ S_q\ (I,t),\ S_r\ (0,t) \Rightarrow S_r\ (I,t+1)$

(ii) $S_p\ (I,t),\ S_q\ (R_i,t),\ S_r\ (0,t) \Rightarrow S_r\ (I,t+1)$ only if $i = 1$ or $i = 2$, and $S_q \rightarrow S_r$

(iii) $S_p\ (I,t),\ S_q\ (R_i,t),\ S_r\ (0,t) \Rightarrow S_r\ (P_1\ P_i,t+1)$ only if $i > 2$, $S_p \longrightarrow S_r$, and $S_q \longrightarrow S_r$

(iv) $S_p\ (I,t),\ S_q\ (P_i,t),\ S_r\ (0,t) \Rightarrow S_r\ (I,t+1)$ only if $i = 1$ or $i = 2$, $S_q \longrightarrow S_r$, and S_q is a sequence set

(v) $S_p\ (I,t),\ S_q\ (P_i,t),\ S_r\ (0,t) \Rightarrow S_r\ (P_1\ P_i,t+1)$ only if $i > 2$, $S_p \longrightarrow$ Sr, $S_q \longrightarrow S_r$, and S_q is a sequence set

(vi) $S_p\ (P/R_i,t),\ S_q\ (P/R_j,t),\ S_r\ (0,t) \Rightarrow S_r\ (P_i\ P_j,t+1)$ only if $i > 2$, $S_p \longrightarrow S_r$, $S_q \longrightarrow S_r$, (and sets exhibiting P at t are sequence set)

(vii) $S_p\ (P/R_i,t),\ S_q\ (P/R_i,t),\ S_r\ (0,t) \Rightarrow S_r\ (P_i,\ t+1)$ only if $i > 2$, $S_p \longrightarrow S_r$, $S_q \longrightarrow S_r$, (and sets exhibiting P at t are sequence set)

Briefly, according to this proposal, each neuronal set is allowed to exhibit multiple activity states. Input units can exhibit several reverberations and can exhibit multiple priming at a time, depending on their history of external stimulation and the amount of priming they receive through connections in the network. In contrast, sequence sets can exhibit ignition, reverberation, or priming, and in addition, they can exhibit multiple additional priming. In activity tables, different activity levels of one set are written on top of each other, the highest activity level appearing in the top line and the lower activity levels in the lines following (see tables in Excursuses E4 and E5).

These formulations include occasional redundancies. For example, (3′iv) follows from (3′ii) and (3′iii) if both principles are applied at the same time and if they are assumed to act independently on different levels of reverberation. Such "additivity" of processing at different levels of activity is assumed from now on. This amounts to considering each neuronal set as a pushdown store with the possibility of storing several hierarchical activity tags that are stored and retrieved independently (Pulvermüller, 1993, 2002). Only the highest activity tags denote "visibility" of the respective activity state and have an immediate influence on ongoing computations.

Note that in the present context activity states of sets, rather than the sets themselves, are called *visible*. What has been called a *visible set* earlier would now be *a set with a visible activity state*.

If an input unit has ignited repeatedly as a result of external input, it exhibits several reverberation states at a time. If the input unit is connected to more than one lexical category representation, more than one of these representations can be activated – coding the information that the repeated element in the input has been classified as member of different lexical categories. To specify inhibitory interactions between competing lexical category representations in the case of multiple reverberation of one input unit, principle (A6) must be extended. The additional assumption proposed next excludes simultaneous preactivity at levels R_1 or P_1 of two category representations connected to one input unit.

(A6′) Inhibition between two lexical category representations α and β connected to an active input unit S (which ignites or reverberates at R_1):

(i) Ignition of S can spread only to sets included in either α or β. The most strongly activated representation wins.
(ii) If one or more sets included in α but not in β ignite at t, then no set included in β and not included in α can exhibit I, P_1, or R_1 at $t+1$.
(iii) If all sets included in α but not in β are reverberating at R_1 at t, then no set included in β and not included in α can exhibit P_1 or R_1 at $t+1$.

Criteria for deciding which of two alternative lexical category representations is more active (A6′i) are discussed in Chapter 10. Axioms (A6′ii) and (A6′iii) would be thought to imply that activity levels P_1 or R_1 of sets in β drop to P_2 or R_2, respectively. Note the modification introduced by (A6′iii). In the case of multiple reverberation of input unit S, it is now possible to have priming effects of reverberation R_1 exerted on α, whereas priming effects of a simultaneous reverberation R_i – at a lower level of R_i, $i > 1$ – can still be effective and affect β.

Intuitively, (A6′i)–(A6′iii) can be attributed to a process of inhibition between lexical category representations that may be "enabled" by an input unit that represents an ambiguous symbol, as detailed earlier.

The simulations presented in Excursuses E4 and E5 show these principles at work on sentences.

12.3 Multiple Center Embedding

Apart from multiple occurrence of individual lexical items, the organization of complements and adjuncts and the representation of lexical category information, a fundamental language problem that can be solved by humans but by no other species, arises in the parsing, analyzing, and understanding of center-embedded sentences. Sentence (7) is an example of how a neuronal grammar consisting of neuronal sets can process multiple center embeddings.

(7) Betty who John who Peter helps loves gets up.

One may question whether this is, indeed, a correct English sentence. One may hold that multiple embeddings of this kind are highly unusual, play no role in conversations, and are considered bad style, even in written language. Therefore, why should a grammar model focusing on the putative brain mechanisms care about them?

The reason is historical. Chomsky (1957, 1963) rejected neuronal finite state automata as devices for describing languages, a main reason being that they fail at processing recursively enumerable languages with multiple center embeddings. Context-free rewriting systems were necessary, he then argued. This led to a paradigm shift, from neuronal (finite-state) to abstract (pushdown) automata. Sentences such as (7) were, so to speak, the flagships of the paradigm shift.

In this context, it is claimed that a neuronal grammar is capable of handling these then crucial strings. Readers who believe that strings such as (7) are ungrammatical may tolerate some elaboration of this issue here for historical reasons. The intention is to show that this string type cannot be used to ground arguments against neuronal models of syntax.

One further extension of the framework is necessary for allowing the network to process center-embedded sentences. According to Braitenberg (1996), some words included in the category of grammatical function words may have a special neuronal representation in that their activation not only implies excitation of a defined neuronal set, but also has an effect of the activity state of the entire brain. An example would be words that prepare the listener to process larger more complex sentence types. Processing of more complex structures may require that activity levels are lowered more than usual to avoid the undesirable consequence of having too much activity and the concomitant interference of activity patterns (cf. Section 5.1).

Words that signal the occurrence of complex string types are, for example, relative pronouns such as *that, who, whom*, or *which* that occur at the start of the subordinate clause. These and other words that, so to speak, change (or "translate") a main sentence into a subordinate sentence have been labeled *translatives* in approaches to dependency grammar (Heringer, 1996; Tesnière, 1959). Because such words warn the listener that there is much more information to come than in an ordinary sentence, it can be postulated that they lead to downward regulation of brain activity. A simple way to model this in a neuronal grammar automaton is to change the global activity level whenever an element of the translative category occurs in the input.

Translatives such as *that* and *who* could therefore have a strong deactivating effect, similar to activation of threshold control by unprimed ignition. It is assumed here that their occurrence in the input activates threshold control regardless of their previous state of activity, and even more strongly than unprimed full ignitions do. The translative-induced activity change is labeled $I^{\wedge\wedge}$, implying double activation of threshold control with the ignition in the context of translatives. Principle (A5) is adjusted to this assumption as follows:

(A5′) Threshold regulation

(i) If a full ignition $I^{\wedge}$ of an input unit S happens at t, then for all reverberating S:
$S\,(R_i, t) \Rightarrow S\,(R_{i+1}, t+1)$

(ii) If there is no ignition or reverberation R_1 at t, then for all reverberating sets S:
$S\,(R_i, t) \Rightarrow S\,(R_{i+1}, t+1)$

(iii) External stimulation of an input unit $S\mathrm{t}$ representing a translative causes double activation $I^{\wedge\wedge}$ of threshold regulation, and for all sets S reverberating at t:
$S\mathrm{t}\,(I^{\wedge\wedge}, t) \Rightarrow S\,(R_{i+2}, t+1)$

Excursus E5 discusses processes in a neuronal grammar circuit induced by a sentence with multiple center embeddings in greater detail.

12.4 Summary, Open Questions, and Outlook

This chapter refines the approach to neuronal grammar proposed in Chapters 10 and 11. The main points relate to a putative neurobiological distinction between obligatory complements of a word and its free adjuncts. The adjunct relationship is proposed to be realized by unidirectional connections, an assumption that is not in conflict with the neuroanatomical reciprocity principle because it is based on unlearning of existing connections. Multiple reverberation was proposed to be possible in neuronal sets. This radical statement is not motivated by empirical data, but the perspectives it offers appear to be a necessary component of any approach to language. There may be alternative candidate neuronal mechanisms for this, but they must first be found. Ignition caused by external input is proposed to lead to an additional activity wave traveling in a word web or input unit. The machinery that may support the processing of center-embedded sentences is discussed in detail, although there is doubt whether this sentence type is of importance in ordinary language use. The ideas and proposals put forward in this chapter are further explored and elaborated in the context of two more simulations detailed in Excursuses E4 and E5.

Considering this approach to grammar in the brain as a whole, general criticisms can be made. Its obvious limitations can be taken as desiderata and questions for future research. For example: The important question of how a neuronal grammar can be learned is addressed only in passing. The postulate is that there is a small set of neuronal principles underlying syntactic knowledge. These principles, or at least relevant aspects of them, are grounded in neurophysiology and neuroanatomy and are, as such, genetically determined, or at least under strong influence of the genetic code. In Chapters 10–12, the postulate is that the neuroscientific principles on which grammar rests include complex aspects of neuronal wiring and function as follows: The specialization of neuron ensembles as sequence sets and word webs, the distinct types of activity states of neuronal sets and their "pushdown dynamics," the criterion for string acceptance and the other principles formulated by principles (A1) and (A2′)–(A6′). Everything beyond these neuronal mechanisms – namely, the connections between word webs and sequence sets and the connection between different sequence sets – are assumed to be the result of associative learning. At this point, it appears one of the most important desiderata to spell out these learning mechanisms in greater detail. Closely connected to this issue is the question of how syntactic category representations could develop and how their mutual inhibition is established. Some preprogrammed information may be relevant here as well. Another open question addresses the mechanisms on which the criteria for string acceptance are based. What kind of neuronal device could check whether the three criteria postulated are all met? These points are relevant for developing further the present proposal.

There is a multitude of important linguistic issues as well that are not addressed here. Among them are questions about the connection between the computation of the structure of a sentence and that of its semantics and about mechanisms underlying the speech acts performed by using sentences. There are open questions about dynamic relationships between word forms and meaning, such as between pronouns and their referents, and about the relationship between the meanings of consecutive sentences. All these issues are not addressed in this book. However, the general perspective on syntactic–semantic interaction the present approach may offer should be sketched briefly.

The proposal in Chapter 4 and elsewhere was that words are represented as distributed neuron webs. The information concerning the form and meaning of a word are processed within one functional system. Because form and meaning representations are tightly linked, the serial-order machinery – what is called *neuronal grammar* – influences both form and meaning representations. The input provided by sequence sets may primarily reach the word form representations of word webs, but because the form and meaning representation of a word are closely connected, the serial-order machinery ultimately exerts an influence on the semantic parts of the word webs as well. The idea is that it is the very connection to the network of sequence detectors that modifies activity in word webs in a way that also has an effect on the semantic parts of the cell assemblies. This could be a neuronal route to the interaction of form-related and semantic information. Again, this must be developed and translated into algorithms and networks.

From a formal linguistic point of view, networks consisting of neuronal sets can accept many of the strings defined by one-sided linear grammars. After all, they are a further development of finite-state networks. It has also been argued that some sets of strings that are outside the reach of finite state grammars (the set of center-embedded strings) and even outside the reach of context-free rewriting systems (certain sentences with subject–verb agreement and particle verbs) can be accepted and processed successfully by neuronal grammar. However, a more precise definition of the sets of strings that can be processed by neuronal grammar may be desirable.

One important issue neglected here is the generation of a sentence: all simulations focus on network dynamics induced in a network by a string in the input. To work out the productive side of neuronal grammar constitutes another entry in the list of to-be-solved tasks.

As mentioned, many syntactic phenomena classic approaches to syntax can handle are not discussed here in the context of neuronal grammar mechanisms. It may be useful, however, to list syntactic phenomena addressed briefly in this chapter that have been addressed in detail on the basis of circuits. This does not imply that the solutions offered are considered optimal or satisfactory.

- Lexical category and subcategory
- Valence
- Dependency between adjacent constituents
- Long-distance dependency
- Distributed word and discontinuous constituent
- Subject–verb agreement
- Adjunct relation
- Multiple word use in a sentence
- Resolution of lexical ambiguity
- Subordination
- Multiple center embedding.

EXCURSUS FOUR

Multiple Reverberation for Resolving Lexical Ambiguity

This excursus illustrates circuits motivated by the proposals discussed in Chapter 12. The abbreviations used here and the representation of activity dynamics in table form are the same as those used in Excursuses E2 and E3.

With the extensions of the grammar circuits proposed in Chapter 12, it now becomes possible to treat sentences in which the same word occurs twice and as member of different lexical categories, such as sentence (1).

(1) Betty switches the switch on.

In Chapter 10, Figures 10.3 and 10.4 are used to sketch a putative neuronal correlate of the syntactic category representations that may be connected to the representation of the word form *switch*. Two of the lexical categories, transitive particle verb, here abbreviated as V, and accusative noun, here abbreviated as N, are relevant for the processing of the syntactically ambiguous word used in sentence (1). The homophonous words and their lexical categories are characterized by the assignment formulas (2) and (3) and the valence formulas (4) and (5).

(2) switch $\leftrightarrow$ V (transitive particle verb)
(3) switch $\leftrightarrow$ N (accusative noun)
(4) V $(p\ /{*}/\ f_1, f_2, f_3)$
(5) N $(p_1, p_2\ /{*}/)$

Figure E4.1 shows the entire network used for sentence processing. The representation of the ambiguous word form is doubled for ease of illustration. This figure is almost identical to Figure E3.1, which dealt with a similar sentence. Table E4.1 presents activity dynamics of the sets involved in processing the ambiguous word and its two lexical categories. The overall network dynamics are very similar to those in the derivation illustrated in

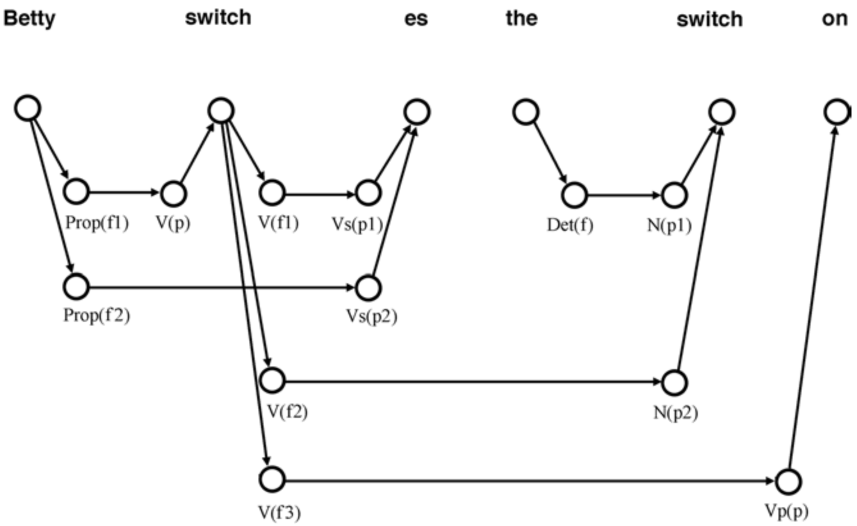

Figure E4.1. Fragment of neuronal grammar that processes and accepts the sentence "Betty switches the switch on." (An equivalent formalism is provided in the text.) Aspects of activation and deactivation dynamics of the network are shown in Table E4.1.

Table E3.1, with the exception that in this case, one item is processed twice. Because the processes are so similar, the specific features of the double processing of the lexically ambiguous word are focused on here.

In contrast to the earlier derivation, Table E4.1 displays only activity states of the input unit representing the repeated element and the relevant sequence sets directly connected to it. An additional difference in the notation must be mentioned: The multiple activity states in one neuronal set are written on top of each other, with the higher activity state always listed above the lower state. Each individual neuronal set now becomes similar to a local pushdown stack that can share activity states with other such stacks. This derivation exploits this for only one set, the word form representation of *switch*. In Excursus E5, the pushdown machinery is used to a greater extent.

There are two sets of sequence sets representing the accusative noun, N, and the transitive particle verb, V. These category representations are assumed to exert inhibition on each other, as specified by (A6′). The effect of inhibition is the channeling of ignitions to only one of the alternative category representation, as specified by (A6′i). This type of channeling of ignitions occurs at time steps 15, 18, 23, and 26 of the simulation. Suppression of reverberation and priming levels caused by ignitions in the representations of competing lexical categories is another important mechanism that becomes effective at time steps 17 and 25.

Table E4.1. Time course of activation and deactivation of word webs and sequence sets of Figure E4.1 during processing of the string "Betty switches the switch on," which is congruent with the network structure. Lines represent time steps. Time steps and input to the network are listed on the left. In each column, activity states of three past-oriented sequence sets (p; to the left of slashes), of a word web (for the word form *switch*), and of three future-oriented sequence sets (f; to the right of slashes) are shown. The two leftmost sequence sets (labeled Det/ and Vt/) make up the lexical category representation of an accusative noun, and the other four sequence sets (Prop/, /Vs, /N4, /Vpart) make up that of a transitive particle verb. Time steps between inputs are chosen so that activity spreads until stable states are reached. Letters indicate ignition (*I*), reverberation (*R*), priming (*P*), and rest (0). *I*^ indicates full unprimed ignition. P_1 and R_1 denote the highest levels of priming and reverberation. Numbers >1 following *P* or *R* imply lower activity ranks, with higher numbers indicating lower activity levels.

	N(p_1) Det/	**N(p_2) Vt/**	**V(p) Prop/**	**/switch/**	**V(f_1) /Vs**	**V(f_2) /N**	**V(f_3) /Vpart**
0	0/	0/	0/	/0/	/0	/0	/0
1 Betty	input unit of Pn "Betty" ignites (full I^)						
2	0/	0/	P_1/	/0/	/0	/0	/0
3	0/	0/	P_1/	/P_1/	/0	/0	/0
4 switch	0/	0/	P_1/	/I/	/0	/0	/0
5	0/	0/	I/	/R_1/	/P_1	**/P_1**	/P_1
6	0/	**P_1/**	R_1/	/R_1/	/P_1	/P_1	/P_1
7 -es	input unit of Vs "-es" ignites (primed I)						
8	0/	P_1/	R_1/	/R_1/	/I	/P_1	/P_1
9	0/	P_1/	R_1/	/I/	/R_1	/P_1	/P_1
10	0/	P_1/	I/	/R_1/	/R_1	/P_1	/P_1
11 the	input unit of "the" ignites (full I^)						
12	P_1/	P_2/	R_2/	/R_2/	/R_2	/P_2	/P_2
13	P_1/	P_2/	R_2/	**/P_1/** **/R_2/**	/R_2	/P_2	/P_2
14 switch	P_1/	P_2/	R_2/	**/I/** **/R_2/**	/R_2	/P_2	/P_2
15	I/	**I/**	R_2/	/R_1/ /R_2/	/R_2	/P_2	/P_2
16	R_1/	R_1/	R_2/	/R_1/ /R_2/	/R_2	**/I**	/P_2
17	R_2/	R_2/	R_2/	/I/ /R_2/	/R_2	/R_1	/P_2
18	R_2/	R_2/	I/	/R_1/ /R_2/	/I	/R_1 /P_2	/P_1 /P_2

	N(p_1) Det/	N(p_2) Vt/	V(p) Prop/	/switch/	V(f_1) /Vs	V(f_2) /N	V(f_3) /Vpart
19	R_2/	P_1/ R_2/	R_1/	/R_1/ /R_2/	/R_1 /P_2	/R_1 /P_2	/P_1 /P_2
20 on	input unit of "on" ignites (primed I)						
21	R_2/	R_2/	R_1/	/R_1/ /R_2/	/R_1 /P_2	/R_1 /P_2	/I /P_2
22	R_2/	R_2/	R_1/	/I/ /R_2/	/R_1 /P_2	/R_1 /P_2	/R_1 /P_2
23	R_2/	R_2/	I/	/R_1/ /R_2/	/I /P_2	**/I**	/R_1 /P_2
24	R_2/	**I/**	R_1/	/R_1/ /R_2/	/R_1 /P_2	/R_1 /P_2	/R_1 /P_2
25	R_1/	R_1/ P_2/	R_2/	/I/ /R_2/	/R_2	/R_2	/R_2
26	I/	R_1/ P_2/	R_2/	/R_1/ /R_2/	/R_2	/R_2	/R_2
27	R_1/	R_1/ P_2/	R_2/	/R_1/ /R_2/	/R_2	/R_2	/R_2
28 DEACTIVATE	0/	0/	0/	/0/	/0	/0	/0

Two of the sequence sets directly connected with the input unit of the ambiguous word form *switch* are also directly connected to each other. The backward-looking sequence detector of the N designed to detect a preceding V are directly connected to the forward-looking sequence detector of V sensitive to a follower N on the right. This leads to a circle in the network, as sketched by (6).

(6) switch ⟶ V (f$_2$) ⟶ N (p$_2$) ⟶ switch

The circle can give rise to reverberation of activity between neuronal sets. It is this loop that explains why, after activating the verb representation together with the word form *switch*, the network later tends to classify it as a noun. At time step 6, N (p$_2$) is being primed through this loop, and at time steps 14–15 and 23–24, ignitions spread through it, notably in different directions.

The derivation from time steps 1–5 is standard. The first occurrence of the word form is classified as a verb, because it has been preceded by a proper name (Prop). The verb suffix (Vs) provides further support for this classification (steps 8–10). The determiner (Det) in the input causes full ignition and therefore reduction of all activity levels. At this point, sequence sets being part of the lexical category representation of N have accumulated more activity [they exhibit priming levels of P_1 and P_2 (average: 1.5), respectively, whereas all activity levels of the competing category are at P_2 or R_2 (average: 2)].

Therefore, the next ignition of the input unit representing the word form *switch* ignites only the sets included in the N representation [according to (A6′i) – see time step 14]. Notably, classification of the second occurrence of the repeated word as an N provides further support for the correctness of the classification of its first occurrence as a verb. This is expressed by the ignition of V (f_2) at step 16, which leads to ignitions of almost all sequence sets included in the verb category representation. This ignition also reduces the activity levels of the competing and reverberating lexical category representation [as stated by (A6′iii)].

After presentation of the final word in the sentence, a wave of ignitions spreads through the network causing synchronization of the already visible and satisfied sets. It is noteworthy that during the spreading process, the hierarchy of reverberation of the two competing lexical category representations is again being changed (time steps 24–25).

This example demonstrates that a network composed of neuronal sets can process sentences in which a word form occurs twice but as a member of different lexical categories. The processing not only includes "accepting" the string, but a spatiotemporal pattern of activity also arises that includes information about the hierarchical organization of the sentence and the assignment of morpheme tokens to lexical categories. Most notable is the dynamic linking of the word-form representation with its alternative lexical category representations in the course of the processing of a sentence that includes an ambiguous word.

One may still ask many further questions, however; for example, about how three or more occurrences of the same word form would be processed. Nevertheless, the proposed framework poses no a priori limit to the number of occurrences of a word, and there appears to be no principled reason why a network composed of neuronal sets should be unable to deal with a construction in which a given word form occurs more than twice. The introduced examples may suffice for illustrating perspectives of the present approach.

EXCURSUS FIVE

Multiple Reverberations and Multiple Center Embeddings

This excursus presents and discusses a very complex sentence, which is actually a word chain whose status as a grammatical sentence may be questioned. The interested reader may nevertheless find it relevant to glance at the processes, because of the prominent role sentences of this type played in the history of language science. The present circuit proves that a string with center embeddings can be processed by neuronal grammar. There is nothing in the neuronal algorithm that would restrict the number of embeddings possible. Clearly, however, such restrictions apply for biological systems.

Consider sentence (1).

(1) Betty who John who Peter helps loves gets up.

Figure E5.1 presents the network representations of the elements of this sentence, word webs and sequence sets, and their mutual connections, whereas Table E5.1 shows the derivation. It becomes obvious from the table that this derivation draws heavily on multiple activation of sequence sets. As in Excursus E4, the table lists multiple activity states of each neuronal set at each time step, with the highest activity state listed at the top. Note that, according to the present proposal, each neuronal set can be considered a store in which multiple entries can be placed on each other. A similar proposal was made by Schnelle (1996b). The little "stacks" of symbols represent the entries in the pushdown memory characterizing each set. The maximum "height" of a stack is three levels.

The algorithmic equivalent of the network in Figure E5.1 follows. The lexical items are assigned to the following categories:

(2) Who $\leftrightarrow$ Tr (translative)

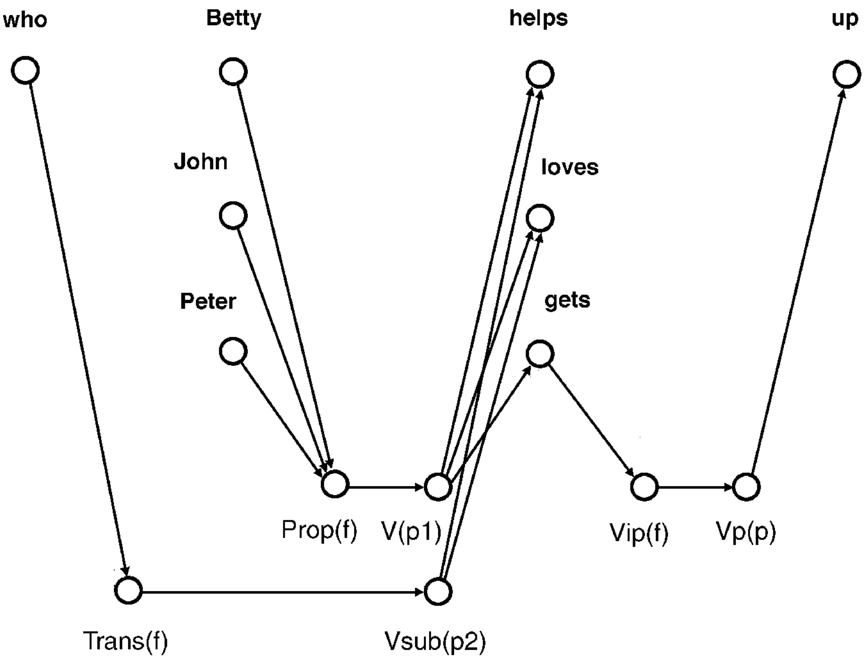

Figure E5.1. Fragment of neuronal grammar that processes and accepts the sentence "Betty who John who Peter helps loves gets up." (An equivalent formalism is provided in the text.) Activation and deactivation dynamics of the network are shown in Table E5.1.

(3) Betty, John, Peter ↔ Prop (nominative proper name)
(4) helps, loves ↔ Vsub (subordinate transitive verb)
(5) gets ↔ Vip (intransitive particle verb)
(6) up ↔ Vpart (verb particle)

With these assignment formulas, a simplification is made compared with the grammars in earlier chapters. Verb suffixes are again treated as parts of the verbs. This is done to shorten the derivation. Previous examples had shown how noun–verb agreement is modeled. The same applies to a more elaborate version of the present grammar.

The lexical categories are characterized by the following valence properties:

(7) Tr (/∗/ f)
(8) Prop (/∗/ f)
(9) Vsub (p_1, p_2 /∗/)
(10) Vip (p /∗/ f)
(11) Vpart (p /∗/)

This proposal implies that verbs in main clauses and in subordinate sentences are attributed to distinct lexical categories. The main clause verb would search for its complements in the past and future, so to speak, whereas a verb in a subordinate clause would restrict search to the past.

Connections between sequence sets are expressed by the following sequence formulas:

(12) Tr (f) $\longrightarrow$ Vsub (p_2)
(13) Prop (f) $\longrightarrow$ Vsub (p_1)
(14) Prop (f) $\longrightarrow$ Vip (p)
(15) Vip (f) $\longrightarrow$ Vpart (p)

Formulas (13) and (14) denote the same connection; sets Vsub (p_1) and Vip (p) are considered identical. Therefore, these formulas could, in principle, be replaced by (16), which then refers to a connection shared between the two alternative representations of the two distinct lexical category representations of verbs appearing in main and subordinate sentences, respectively. Also, the sequence detector V(p) is shared by all verb representations.

(16) Prop (f) $\longrightarrow$ V (p)

Note again that in the present grammar fragment, a distinction is made between ordinary verbs in main clauses and verbs in subordinate sentences (Vsub), which are assumed to take translatives as their complements. Translatives take the place of the accusative noun as the verb complement. However, in contrast to ordinary accusative nouns that must appear to the right of the verb, translatives are placed to the left of transitive verbs. This motivates the postulate of a distinct set of sequence sets for the new lexical category Vsub.

To model this, some grammars introduce transformations, movements, or other sophisticated and costly procedures. In the proposed neuronal grammars, the solution is different. Each transitive verb is supplied with two distinct sets of sequence features. The "standard" form (10) is characterized by nominative and accusative complements occurring to the left and right of the verb, respectively, and the probably less-frequent subordinate form (9), in which both complements occur to the left and one complement is a translative.

Postulating a small number of additional sets of sequence features for lexical categories appears less costly than introducing new operations, such as transformations or movements (see Chapter 10, Section 10.6.4). After all, the number of sentence forms consisting of the same lexical categories but differing in their order is small in English. Introduction of distinct sets of sequence features for verbs occurring in sentence forms such as active,

Table E5.1. Time course of activation and deactivation of word webs and sequence sets of Figure E5.1 during processing of the string "Betty who John who Peter helps loves gets up," which is congruent with the network structure. Lines represent time steps. Time steps and input to the network are listed on the left. In each column, activity states of one or more optional past-oriented sequence sets (p; to the left of slashes), of a word web (between slashes), and of optional future-oriented sequence sets (f; to the right of slashes) are shown. Time steps between inputs are chosen so that activity spreads until stable states are reached. Letters indicate ignition (*I*), reverberation (*R*), priming (*P*), and rest (0). *I*^ indicates full unprimed ignition, and *I*^^ an ignition causing strong threshold regulation. P_1 and R_1 denote the highest activity level. Numbers >1 following *P* or *R* mean activity ranks, with higher numbers indicating lower levels.

	/Who/f	**/Betty/f**	**/John/f**	**/Peter/f**	**p_1,p_2/Helps/**	**p_1,p_2/Loves/**	**p/Gets/f**	**l/Up/**
0	/0/0	/0/0	/0/0	/0/0	0,0/0/	0,0/0/	0/0/0	0/0/
1 Betty	/0/0	/I^/0	/0/0	/0/0	0,0/0/	0,0/0/	0/0/0	0/0/
2	/0/0	$/R_1/P_1$	$/0/P_1$	$/0/P_1$	0,0/0/	0,0/0/	0/0/0	0/0/
3	/0/0	$/R_1/P_1$	$/0/P_1$	$/0/P_1$	P_1,0/0/	P_1,0/0/	P_1/0/0	0/0/
4	/0/0	$/R_1/P_1$	$/0/P_1$	$/0/P_1$	P_1,0/0/	P_1,0/0/	P_1/P_1/0	0/0/
5 who	/I^^/0	$/R_3/P_3$	$/0/P_3$	$/0/P_3$	P_3,0/0/	P_3,0/0/	P_3/P_3/0	0/0/
6	$/R_1/P_1$	$/R_3/P_3$	$/0/P_3$	$/0/P_3$	P_3,0/0/	P_3,0/0/	P_3/P_3/0	0/0/
7	$/R_1/P_1$	$/R_3/P_3$	$/0/P_3$	$/0/P_3$	P_3,P_1/0/	P_3,P_1/0/	P_3/P_3/0	0/0/
8	$/R_1/P_1$	$/R_3/P_3$	$/0/P_3$	$/0/P_3$	P_3,P_1/P_2/	P_3,P_1/P_2/	P_3/P_3/0	0/0/
9 John	$/R_2/P_2$	$/R_4/P_4$	/I^/P_4	$/0/P_4$	P_4,P_2/P_3/	P_4,P_2/P_3/	P_4/P_4/0	0/0/
10	$/R_2/P_2$	$/R_4/P_1$ $/P_4$	$/R_1/P_1$ $/P_4$	$/0/P_1$ $/P_4$	P_4,P_2/P_3/	P_4,P_2/P_3/	P_4/P_4/0	0/0/
11	$/R_2/P_2$	$/R_4/P_1$ $/P_4$	$/R_1/P_1$ $/P_4$	$/0/P_1$ $/P_4$	P_1,P_2/P_3/ P_4, /	P_1,P_2/P_3/ P_4, /	P_1/P_4/0 P_4/	0/0/

12	$/R_2/P_2$	$/R_4/P_1$ $/P_4$	$/R_1/P_1$ $/P_4$	$/0/P_1$ $/P_4$	$P_1,P_2/P_1/$ $P_4, /$	$P_1,P_2/P_1/$ $P_4, /$	$P_1/P_1/0$ $P_4/$	0/0/
13 who	/I^^/P_4 $/R_4/$	$/R_6/P_3$ $/P_6$	$/R_3/P_3$ $/P_6$	$/0/P_3$ $/P_6$	$P_3,P_4/P_3/$ $P_6, /$	$P_3,P_4/P_3/$ $P_6, /$	$P_3/P_3/0$ $P_6/P_6/$	0/0/
14	$/R_1/P_1$ $/R_4/P_4$	$/R_6/P_3$ $/P_6$	$/R_3/P_3$ $/P_6$	$/0/P_3$ $/P_6$	$P_3,P_4/P_3/$ $P_6, /$	$P_3,P_4/P_3/$ $P_6, /$	$P_3/P_3/0$ $P_6/P_6/$	0/0/
15	$/R_1/P_1$ $/R_4/P_4$	$/R_6/P_3$ $/P_6$	$/R_3/P_3$ $/P_6$	$/0/P_3$ $/P_6$	$P_3,P_1/P_3/$ $P_6,P_4/$	$P_3,P_1/P_3/$ $P_6,P_4/$	$P_3/P_3/0$ $P_6/P_6/$	0/0/
16	$/R_1/P_1$ $/R_4/P_4$	$/R_6/P_3$ $/P_6$	$/R_3/P_3$ $/P_6$	$/0/P_3$ $/P_6$	$P_3,P_1/P_2/$ $P_6,P_4/P_5/$	$P_3,P_1/P_2/$ $P_6,P_4/P_5$	$P_3/P_3/0$ $P_6/P_6/$	0/0/
17 Peter	$/R_2/P_2$ $/R_4/P_5$	$/R_7/P_4$ $/P_7$	$/R_4/P_4$ $/P_7$	/I^/P_4 $/P_7$	$P_4,P_2/P_3/$ $P_7,P_5/P_6/$	$P_4,P_2/P_3/$ $P_7,P_5/P_6$	$P_4/P_4/0$ $P_7/P_7/$	0/0/
18	$/R_2/P_2$ $/R_5/P_5$	$/R_7/P_1$ $/P_4$ $/P_7$	$/R_4/P_1$ $/P_4$ $/P_7$	$/R_1/P_1$ $/P_4$ $/P_7$	$P_4,P_2/P_3/$ $P_7,P_5/P_6/$	$P_4,P_2/P_3/$ $P_7,P_5/P_6/$	$P_4/P_4/0$ $P_7/P_7/$	0/0/
19	$/R_2/P_2$ $/R_5/P_5/$	$/R_7/P_1$ $/P_4$ $/P_7$	$/R_4/P_1$ $/P_4$ $/P_7$	$/R_1/P_1$ $/P_4$ $/P_7$	$P_1,P_2/P_3/$ $P_4,P_5/P_6$ $P_7, /$	$P_1,P_2/P_3/$ $P_4,P_5/P_6/$ $P_7, /$	$P_1/P_4/0$ $P_4/P_7/$ $P_7/$	0/0/
20	$/R_2/P_2$ $/R_5/P_5$	$/R_7/P_1$ $/P_4$ $/P_7$	$/R_4/P_1$ $/P_4$ $/P_7$	$/R_1/P_1$ $/P_4$ $/P_7$	$P_1,P_2/P_1/$ $P_4,P_5/P_4$ $P_7, /$	$P_1,P_2/P_1/$ $P_4,P_5/P_4/$ $P_7, /$	$P_1/P_1/0$ $P_4/P_4/$ $P_7/P_7/$	0/0/
21 helps	$/R_2/P_2$ $/R_5/P_5$	$/R_7/P_1$ $/P_4$ $/P_7$	$/R_4/P_1$ $/P_4$ $/P_7$	$/R_1/P_1$ $/P_4$ $/P_7$	$P_1,P_2/I$ / $P_4,P_5/P_4/$ $P_7, /$	$P_1,P_2/P_1/$ $P_4,P_5/P_4/$ $P_7, /$	$P_1/P_1/0$ $P_4/P_4/$ $P_7/P_7/$	0/0/

(continued)

Table E5.1 (*continued*)

	/Who/f	/Betty/f	/John/f	/Peter/f	p_1,p_2/Helps/	p_1,p_2/Loves/	p/Gets/f	I/Up/
22	$/R_2/P_2$ $/R_5/P_5$	$/R_7/P_1$ $/P_4$ $/P_7$	$/R_4/P_1$ $/P_4$ $/P_7$	$/R_1/P_1$ $/P_4$ $/P_7$	$I,I\ /R_1/$ $P_4,P_5/P_4/$ $P_7,\ /$	$I,I\ /P_1/$ $P_4,P_5/P_4/$ $P_7,\ /$	$I\ /P_1/0$ $P_4/P_4/$ $P_7/P_7/$	0/0/
23	$/R_2/I$ $/R_5/P_5$	$/R_7/I$ $/P_4$ $/P_7$	$/R_4/I$ $/P_4$ $/P_7$	$/R_1/I$ $/P_4$ $/P_7$	$R_1,R_1/R_1/$ $P_4,P_5/P_4/$ $P_7,\ /$	$R_1,R_1/P_1/$ $P_4,P_5/P_4/$ $P_7,\ /$	$R_1/P_1/0$ $P_4/P_4/$ $P_7/P_7/$	0/0/
24	$/I\ /R_1$ $/R_5/P_5$	$/R_7/R_1$ $/P_4$ $/P_7$	$/R_4/R_1$ $/P_4$ $/P_7$	$/I\ /R_1$ $/P_4$ $/P_7$	$R_1,R_1/R_1/$ $P_4,P_5/P_4/$ $P_7,\ /$	$R_1,R_1/P_1/$ $P_4,P_5/P_4/$ $P_7,\ /$	$R_1/P_1/0$ $P_4/P_4/$ $P_7/P_7/$	0/0/
25	**$/R_1/R_1$** $/R_5/P_5$	$/R_7/R_1$ $/P_4$ $/P_7$	$/R_4/R_1$ $/P_4$ $/P_7$	**$/R_1/R_1$** $P_4,P_5/P_4/$ $/P_7$	**$R_1,R_1/R_1/$** $P_4,P_5/P_4/$ $P_7,\ /$	$R_1,R_1/P_1/$ $P_4/P_4/$ $P_7,\ /$	$R_1/P_1/0$ $P_7/P_7/$	0/0/
26 DEACTIVATE	$/R_5/P_5$	$/R_7/P_4$ $/P_7$	$/R_4/P_4$ $/P_7$	$/0/P_4$ $/P_7$	$P_4,P_5/P_4/$ $P_7,\ /$	$P_4,P_5/P_4/$ $P_7,\ /$	$P_4/P_4/0$ $P_7/P_7/$	0/0/
27 LEVELADJUST	$/R_2/P_2$	$/R_4/P_1$ $/P_4$	$/R_1/R_1$ $/P_4$	$/0/R_1$ $/P_4$	$P_1,P_2/P_1/$ $P_4,\ /$	$P_1,P_2/P_1/$ $P_4,\ /$	$P_1/P_1/0$ $P_4/P_4/$	0/0/
28 loves	$/R_2/P_2$	$/R_4/R_1$ $/P_4$	$/R_1/R_1$ $/P_4$	$/0/R_1$ $/P_4$	$P_1,P_2/P_1/$ $P_4,\ /$	$P_1,P_2/I/$ $P_4,\ /$	$P_1/P_1/0$ $P_4/P_4/$	0/0/
29	$/R_2/P_2$	$/R_4/R_1$ $/P_4$	$/R_1/R_1$ $/P_4$	$/0/R_1$ $/P_4$	$I,I\ /P_1/$ $P_4,\ /$	$I,I\ /R_1/$ $P_4,\ /$	$I\ /P_1/0$ $P_4/P_4/$	0/0/
30	$/R_2/I$	$/R_4/I$ $/P_4$	$/R_1/I$ $/P_4$	$/0/I$ $/P_4$	$R_1,R_1/P_1/$ $P_4,\ /$	$R_1,R_1/R_1/$ $P_4,\ /$	$R_1/P_1/0$ $P_4/P_4/$	0/0/
31	$/I/R_1$	$/R_4/R_1$ $/P_4$	$/I\ /R_1$ $/P_4$	$/0/R_1$ $/P_4$	$R_1,R_1/P_1/$ $P_4,\ /$	$R_1,R_1/R_1/$ $P_4,\ /$	$R_1/P_1/0$ $P_4/P_4/$	0/0/

32	**$/R_1/R_1$**	$/R_4/R_1$	**$/R_1/R_1$**	$/0/R_1$	$R_1,R_1/P_1/$	**$R_1,R_1/R_1/$**	$R_1/P_1/0$	$0/0/$
		$/P_4$	$/P_4$	$/P_4$	$P_4, /$	$P_4, /$	$P_4/P_4/$	
33 DEACTIVATE	$/0/0$	$/R_4/P_4$	$/0/P_4$	$/0/P_4$	$P_4,0/0/$	$P_4,0/0/$	$P_4/P_4/0$	$0/0/$
34 LEVELADJUST	$/0/0$	$/R_1/P_1$	$/0/P_1$	$/0/P_1$	$P_1,0/0/$	$P_1,0/0/$	$P_1/P_1/0$	$0/0/$
35 gets	$/0/0$	$/R_1/P_1$	$/0/P_1$	$/0/P_1$	$P_1,0/0/$	$P_1,0/0/$	$P_1/I /0$	$0/0/$
36	$/0/0$	$/R_1/P_1$	$/0/P_1$	$/0/P_1$	$I,0/0/$	$I,0/0/$	$I /R_1/P_1$	$P_1/0/$
37	$/0/0$	$/R_1/I$	$/0/I$	$/0/I$	$R_1,0/0/$	$R_1,0/0/$	$R_1/R_1/P_1$	$P_1/P_1/$
38	$/0/0$	$/I /R_1$	$/0/I$	$/0/I$	$R_1,0/0/$	$R_1,0/0/$	$R_1/R_1/P_1$	$P_1/P_1/$
39 up	$/0/0$	$/R_1/R_1$	$/0/R_1$	$/0/R_1$	$R_1,0/0/$	$R_1,0/0/$	$R_1/R_1/P_1$	$P_1/I/$
40	$/0/0$	$/R_1/R_1$	$/0/R_1$	$/0/R_1$	$R_1,0/0/$	$R_1,0/0/$	$R_1/R_1/P_1$	$I/R_1/$
41	$/0/0$	$/R_1/R_1$	$/0/R_1$	$/0/R_1$	$R_1,0/0/$	$R_1,0/0/$	$R_1/R_1/I$	$R_1/R_1/$
42	$/0/0$	$/R_1/R_1$	$/0/R_1$	$/0/R_1$	$R_1,0/0/$	$R_1,0/0/$	$R_1/I /R_1$	$R_1/R_1/$
43	$/0/0$	$/R_1/R_1$	$/0/R_1$	$/0/R_1$	$I,0/0/$	$I,0/0/$	$I /R_1/R_1$	$R_1/R_1/$
44	$/0/0$	$/R_1/I$	$/0/I$	$/0/I$	$R_1,0/0/$	$R_1,0/0/$	$R_1/R_1/R_1$	$R_1/R_1/$
45	$/0/0$	$/I /R_1$	$/0/R_1$	$/0/R_1$	$R_1,0/0/$	$R_1,0/0/$	$R_1/R_1/R_1$	$R_1/R_1/$
46	$/0/0$	**$/R_1/R_1$**	$/0/R_1$	$/0/R_1$	$R_1,0/0/$	$R_1,0/0/$	**$R_1/R_1/R_1$**	**$R_1/R_1/$**
47 DEACTIVATE	$/0/0$	$/0/0$	$/0/0$	$/0/0$	$0,0/0/$	$0,0/0/$	$0/0/0$	$0/0/$

passive, question, and relative clause therefore do not lead to too heavy a burden on the grammar machinery. It is also important to recall that introducing new sets of sequence sets is not a problem in the present framework. The development and definition of a system of lexical categories along these lines is left for future research.

The possible brain basis of free adjuncts of a sentence constituent is addressed in Section 12.1. In the present example sentence, the translatives are adjuncts that can be freely added to any noun. Thus, their respective neuronal sets should not exhibit a strong reciprocal connection with the neuronal set representing the word whose adjunct they can be considered. The sets realizing noun and transitive adjunct would connect via weak unidirectional links. These allow for contributing to the reactivation of the mother noun after a subordinated sentence has been processed. The present simulation does not cover adjunct connections because their introduction necessitates further elaboration of the algorithm and table. Because in the present formulation, reactivation of the mother is possible on the basis of threshold regulation [axiom (AS′)], such elaboration appears avoidable here.

Example sentence (1) exhibits a feature discussed in Section 12.2, the multiple occurrence of a word form. In this case, the translative is used twice in the course of the analysis of a sentence. In the example simulation shown in Table E5.1, its double use is reflected in the double activation of its neuronal set starting at time step 13 and ending at time step 25. Because in this simulation the pushdown mechanism separates the processing of the different levels of embedding of the sentence, one of the transitive occurrences – or *tokens* is processed together with the first embedded sentence and the other occurrence with the innermost sentence embedded in both others. This is another illustration of how multiple occurrences of the same item can be processed in a neuronal grammar. Note the dynamical grouping of active representations into sets of sets that interact.

Further crucial steps in the derivation displayed in Table E5.1, the network's processing of a sentence with multiple center embeddings, are as follows: Ignitions of input units of translatives occur at time steps 5 and 13. As a consequence, there is double activation of threshold control, and therefore all "visible" activity states of other sets in the network become "invisible" – that is, their activity levels are reduced by two levels. At time step 5, reduction is to levels R_3 and P_3, and at time step 13, even levels become as low as R_6 and P_6. In other words, the information about the elements that were most active in the device's memory is now being "pushed down" to a lower level of activity, where the activity states are no longer "visible." Once again, this is analogous to the operation of a pushdown mechanism.

According to the modified proposal in Chapter 12, each neuronal set can be considered to include independent pushdown memories.

Because ignitions can be transmitted only through the network to visible sets, or, more precisely, to visible activity levels of sets, ignitions can only affect the sets reverberating and primed at the highest levels of R_1, P_1, R_2, and P_2 (see Chapter 10 for discussion). After all the input units of one of the three embedded sentences have been activated, the respective neuronal elements whose states are visible are synchronized and satisfied. This becomes effective separately for the three sentences embedded into each other, at time steps 25, 32, and 46, respectively. In the table, the relevant activity levels at these time steps are highlighted in bold print. After separate synchronization of each of the three sentences, the respective synchronized parts of the complex sentence, that is, the embedded clauses, are deactivated sequentially (time steps 26, 33, and 47). The three sentences are therefore processed separately, as if they were not embedded into each other, but appeared in an ordinary manner one after the other instead. Note, however, that this analysis leaves open the question of coreference relations between nouns and translatives.

After deactivation of the representations of embedded sentences, several sets reverberate or are primed at low levels of R or P. At this point, the threshold machinery (A5′) must increase activity levels so that the most active sets become visible. As already mentioned, however, strictly speaking, this adjustment of levels is redundant because according to the present assumptions the highest activity levels are visible, regardless of what their absolute values actually are (cf. Chapter 10). If the most strongly activated sets are deactivated, the sets with the next lower levels of activity automatically become visible. Thus, the changes at time steps 27 and 34 are simply the result of relabeling of activity levels rather than the result of real activity changes. Nevertheless, as emphasized in Section 5.1, from a neurobiological perspective it appears necessary to postulate a mechanism for the regulation of activity levels of neuronal elements by both decreasing and increasing activity. Also, this appears advantageous for yielding activity values within the range of optimal functioning. For these reasons, the assumption of a threshold regulation device that can provide both global activation and deactivation is maintained, although one aspect of it may be considered redundant in the present derivations. In real neuronal structures, the regulation device may have the effect that the visible activity levels are always at about the same level.

Several problems cannot be solved by the present network and extensions are necessary. As earlier examples of grammar circuits, the present grammar fragment again produces some false-positive errors. This may be improved by more fine-grained adjustments. The transitives would, according to the present concept, provide the main mechanism for separating the processing of the embedded constructions. However, there is the problem of disambiguation of translatives, because, for example, the word form *who* can be used not only as translative connecting a superordinate to a subordinate sentence but also as an interrogative pronoun in *who* questions. This may not appear to be a difficult problem, and the neuronal connections modeling the adjunct relation and the priming they support may offer perspectives on solving it. Yet an even worse problem appears to be that overt translatives are not obligatory in English, as in string (17).

(17) The woman John loves gets up.

For these cases, a mechanism must be specified that does not rely on overt translatives. One possibility is to postulate that in this case, sequence detectors strongly activate threshold regulation instead of individual words. A further open question concerns the interaction of the proposed serial-order machinery with the semantic representations of words. These and numerous other aspects must be addressed in the future (Crespi-Reghizzi et al., 2001).

After considering this example, the following general concluding remark appears appropriate. Neuronal grammar can accept strings, including center embeddings, by grouping embedded phrases into synchronized assemblies of neuronal sets. This provides a basic structural analysis of complex sentences. It is important that there is, in the present grammar, no upper limit for the number of embeddings or the number of elements assigned to any given lexical category whose representation is being activated multiply, in principle without upper limit for the number of reverberations. Thus, there is the possibility of producing longer and longer strings of lower and lower levels of embedding, and new lexical elements can easily be added to the grammar fragment. There is no theoretical reason for excluding such excessive use of the pushdown machinery. The only reasons are empirical in nature.

From this excursus, it appears relevant to keep in mind that neuronal grammar can accept at least certain subsets of strings defined by context-free rewriting systems. It appears that the flagship sentence structure from the linguistic wars against neuron-based finite state grammars can travel calmly in the light of neuroscientific knowledge and theorizing.

CHAPTER THIRTEEN

Neurophysiology of Syntax

Models are only as good as the tools available for testing them. Is there any perspective on testing the proposals about syntactic brain processes discussed to this point?

13.1 Making Predictions

One obvious prediction of the neuronal grammar framework is that the cortical processes following a sentence in the input are of different types. Such process types include the following:

- Ignitions of word webs
- Ignitions of sequence sets
- Downward or upward regulation of activity by threshold control
- Reverberation and priming of neuronal sets

It may be possible to investigate these putative processes using neurophysiological recording techniques. Because the relevant activation and deactivation processes likely occur within tenths or hundredths of a second, it is advisable to use fast neuroimaging techniques in the investigation of serial-order information processing.

At first glance, one straightforward way of transforming any of the mentioned simulations into predictions about neurophysiological processes appears to be by calculating sum activity values at any time step of a simulation, and by thereby obtaining a series of activity values.

However, this strategy is not so straightforward as it appears at first glance because several parameters must be chosen beforehand. The relative weighting of the activity states is one set of parameters that appears to distinguish between pronounced activity changes – those induced by ignitions and regulation processes – and smaller ones – as, for example, those caused by the spreading of priming or reverberation. As an additional set of parameters, the time constants of individual processes must be specified. The time constants of the postulated processes, such as an ignition and the spreading of priming, must be chosen. Furthermore, in Excursuses E4 and E5, full ignitions and the subsequent regulation of cortical activity contracted into one time step. This is reasonable to avoid lengthy simulations, but because the regulation process is the consequence of the strong activation process, it is necessary to separate the two processes within a model designed to make exact predictions on the time course of cortical activation and deactivation. Predictions from the simulations cannot be 1 : 1 conversions.

13.2 Neuronal Grammar and Syntax-Related Brain Potentials

A principled prediction made by the model is that after a word is present in the input, the word web ignites before its connected sequence sets. However, this prediction is not easy to test. Because preactivated sequence detectors can give rise to the priming of word webs, the same word appearing in different contexts not only activates different sets of sequence sets, but the ignition of its own word web may also differ as a function of context. It is of utmost importance here to choose appropriate baseline conditions for comparison.

A further clear prediction is that syntactically well-formed and ungrammatical word strings elicit differential brain responses. The processes induced in circuits of neuronal grammar by syntactically well-formed (congruent) sentences can be compared with the effect of syntactic violations in (incongruent) strings. As a first step, this can be done purely on the basis of the model simulations. Comparison of these processes (see Excursus E2) shows that the physiological differences between grammatical and ungrammatical strings should be three-fold. A situation is considered in which one particular word or morpheme fits in syntactic context 1, but yields an ungrammatical string in context 2.

The first postulate would be that different types of ignitions should occur: primed ignition in the case of the congruent sentence, but full unprimed ignition after the item has been placed incorrectly. The activation of the unprimed word web should give rise to a more substantial activity increase compared to the primed ignition in the grammatical context. Therefore, there should be evidence of stronger activity increase in the cortex when a word is placed in a context in which it is syntactically incorrect compared to a situation in which it is positioned in a grammatically correct context. This difference should be apparent early, shortly after the information about the new word has reached the brain.

Neurophysiological studies using electroencephalography (EEG) show that certain syntactically deviant word strings elicit an early negative event-related potential (Neville, Nicol, Barss, Forster, & Garrett, 1991). This brain response has frequently been found to be maximal over the left hemisphere dominant for language. It can be elicited by the incorrectly placed word *of* in sentence (1), but other syntactically incorrect strings may also produce it.

(1) The scientist criticized Max's of proof the theorem.

The brain response is sometimes called the *early left-anterior negativity*. In this framework, it would be interpreted as a physiological indicator of the full ignition of the word web of a word placed incorrectly. The early negativity is less pronounced or absent if words are presented in a grammatical sentence. The effect could be reproduced repeatedly for different types of syntactic violations, and with some variation of the cortical topography and latency of the brain response (Friederici, 1997; Friederici & Mecklinger, 1996; Gunter, Friederici, & Schriefers, 2000; Hahne & Friederici, 1999). An early left-anterior negativity was also reported with incorrectly placed inflectional morphemes as well (Weyerts et al., 1997).

The second postulate differential brain responses elicited by grammatical and ungrammatical strings is as follows: The threshold control mechanism becomes active only after full ignition. Thus, activity decrease would be predicted after a strong initial activation process when a syntactically incorrectly placed word or morpheme occurs. Such pronounced activity decrease should be absent when a grammatical word string is being processed.

There is a second neurophysiological brain response that distinguishes syntactically well-formed strings from deviant strings. It is a positive component of the event-related potential and is maximal over the back of the head. Its latency is in the range of 600 ms after onset of the critical word, and is therefore called the *P600 component* or the *syntactic positive shift* (Hagoort, Brown, & Groothusen, 1993; Osterhout & Holcomb, 1992). When an incorrectly placed word is being processed, the late positive shift can appear after an earlier negative shift (Friederici & Mecklinger, 1996; Hahne & Friederici, 1999; Osterhout, 1997).

The suggestion derived from the neuronal grammar model is that the more positive electrical brain response to syntactic anomalies reflects the action of the mechanism that reduces the level of cortical activity after a strong activation process (full ignition of a word) has taken place.

The third postulate neuronal grammar allows for concerns the activation of sequence detectors and the spreading wave of activity following a word that matches its syntactic context. This should again lead to stronger

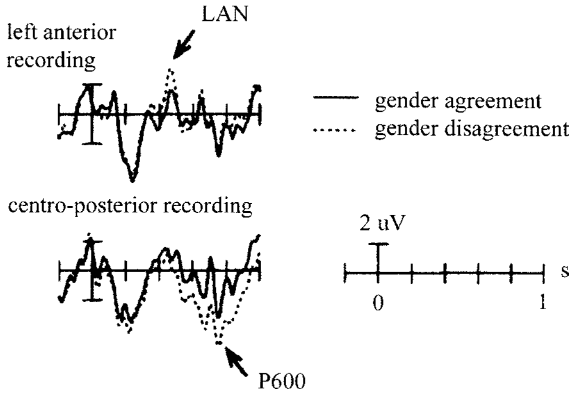

Figure 13.1. Gender violations elicited both an early left-anterior negativity and a late positive shift of the brain potential compared to sentences in which no such violation was present. Example stimuli are the German sentences "Sie bereiste *das* Land" and "Sie bereiste *den* Land," in which the determiner in italics is syntactically anomalous. (Example in English: "She travels the land," in which the article is either in the correct neuter case or in an incorrect case, masculine.) The dotted lines show responses to violations, the bold lines responses to the grammatical sentences. Reprinted with permission from Gunter, T. C., Friederici, A. D., & Schriefers, H. (2000). Syntactic gender and semantic expectancy: ERPs reveal early autonomy and late interaction. *Journal of Cognitive Neuroscience, 12*(4), 556–68.

activation processes when a grammatical string is being processed in comparison to the processing of a string including a syntactic violation. At this point, there does not appear to be evidence for this claim. It should, in principle, be possible to find evidence for this, because if the model is correct, the backward spreading of activity would be a longer lasting process – in particular, if the words are placed in the context of an elaborate sentence.

Concern may arise regarding the latencies at which these syntactically related brain responses have been found. Responses that occur more than one-half second after the information about a word is present in the input may appear unlikely to reflect the earliest aspects of language processes. After 500 ms, it is, in many cases, already possible to respond to a stimulus word, and a brain response occurring at such long latencies is not likely to reflect the processes at the heart of the language machinery (cf. Lehmann & Skrandies, 1984; Pulvermüller, 1999b; Skrandies, 1998).

There are, however, at least two responses to this concern. First, the maximal response can sometimes be seen late, but this does not exclude the possibility that earlier differences are also present. This becomes apparent from Figure 13.1, in which the late positivity may appear to onset substantially before its peak. Second, it is possible that the use of new recording techniques makes it possible to record much earlier onsets of the relevant differences. Many early brain responses are smaller than the late responses appearing approximately one-half second after the stimulus. Because many early responses are small, they are more likely to be masked by noise. For recording earlier dynamics related to the processing of words and serial-order information, it may be crucial to introduce techniques that allow for improving the signal-to-noise ratio. This may make earlier brain dynamics of syntactic processing more visible.

This brief chapter is intended to show that what could appear as a purely theoretical proposal of a neuronal grammar can actually be connected to neurophysiological data. The early syntactic negativity may reflect the unprimed full ignition of the neuronal set of a word placed in a syntactically deviant environment. The late positive shift may reflect activation of the threshold regulation device caused by the substantial activity increase produced by the unprimed ignition. Further important neurophysiological phenomena realted to grammatical processing (e.g., Kluender & Kutas, 1993; Müller, King, & Kutas, 1997) are still awaiting models at the neuronal level that may provide tentative neuroscientific explanations for them. Clearly, future research and physiological research must show which aspects of the model presented in Chapters 10–12 can be maintained and which must be replaced. Predictions appear obvious, and tests are available.

CHAPTER FOURTEEN

Linguistics and the Brain

Linguistics is the study of language. *Language* is a system of brain circuits. To qualify the latter statement, one may cite the father of modern linguistics, Ferdinand de Saussure, who claimed that language (i.e., the language system, or *langue*) is a “concrete natural object seated in the brain” (de Saussure, 1916). If linguistics is the study of language and language is in one sense a system of brain circuits, one would expect linguists to be open to the study of brain circuits. However, the dominating view in linguistics appears to be that language theories must be formulated in an abstract manner, not in terms of neuron circuits. Section 14.1 asks why linguists favor abstract rather than neuron-based formulations of language mechanisms. Section 14.2 discusses a few thoughts about how an abstract theory of language may profit from a brain basis.

14.1 Why Are Linguistic Theories Abstract?

As mentioned, the dominating view in linguistics is that language theories must be formulated in an abstract way. This could be a trivial claim because it is clear that every scientific theory must include abstract concepts. However, this is not the point. The point is that *abstract* in this context excludes explicit reference to the organic basis of the processes described in an abstract fashion. Linguistic theory is abstract in the sense that it does not refer to neurons. Why is this so?

For a scientist, this may be difficult to understand. An astronomer working on star formation would probably be open to discussing molecule clouds that can be inferred from the recordings performed (e.g., with a radio telescope). The “linguistic mentality,” so to speak, transformed to astronomy would result in a scholar who studies stars but refuses to speak about their component substances and driving forces. The scholar may claim that stars should be discussed only in terms of abstract concepts, not in terms of gases and their temperature.

The position that linguistic theories must be abstract appears to be in need of justification, which could be either a principled statement or an excuse.

One may argue that it is unreasonable to specify neuron circuits of language – as there was no hope for a nineteenth-century astronomer to find out about the physical processes occurring in the center of the sun that cause emission of light. One may posit that there is simply not enough knowledge about brain processes – or, 100 years ago, about the interior of the sun. However, given the immense knowledge about how the brain processes language accumulated in the last 20 years or so, it appears more likely that relevant knowledge is already available for clarifying mechanisms of language in brain terms. At least, the tools appear to be available for obtaining the crucial knowledge about the underlying physiological processes, given there are theoretically crucial predictions.

What is necessary, then, are ideas about how to connect the level of language description to that of the description of neurons. Piling up more neurophysiological and imaging data may not help much in this enterprise. Empirical facts do not by themselves form a theory about the generation of sunlight or language. Theoretical work is required in the first place. The theoretical efforts can lead to the generation of predictions that can be addressed in crucial experiments. Lack of empirical data is never a very good excuse for postponing the necessary theoretical labor.

A reasonable excuse for not addressing language in brain terms may take the following possibilities into consideration. Linguists may have difficulty understanding the language of neuroscientists, and, conversely, neuroscientists may have difficulty understanding linguistic terminology. After all, the distance between linguistics and neuroscience is not smaller than that between physics and chemistry. Given that such mutual comprehension difficulty is relevant, the important problem may be a problem of translation (Schnelle, 1996a, 1996c). Therefore, it may appear relevant to provide translations between the language of linguistic algorithms and that of nerve cells, their connections, and their activity states. Again, a good deal of theoretical work is required.

In one publication, Chomsky (2000) offers another reason for not talking about neuronal mechanisms of language. “It may well be that the relevant elements and principles of brain structure have yet to be discovered” (p. 25). However, brain-theoretical concepts referred to by expressions such as *cell assembly*, *synfire chain*, and *memory cell* are available, and some of the potentially relevant principles of brain structure and function are well understood and have been available for theorizing about language for some time, although language researchers have only very recently been taking advantage of them. However, Chomsky is partly correct. Developing new concepts may be necessary when theorizing about brain mechanisms of language. The main effort undertaken in this book is to propose and sharpen concepts – for example, those labeled *neuronal set* and *multiple reverberation*.

In summary, it seems there is no good reason why linguistic theories should necessarily be abstract, and why they should not be formulated in concrete terms – for example, in terms of neurons. Rather, this appears to be a crucial linguistic task of the future.

14.2 How May Linguistic Theory Profit from a Brain Basis?

A question frequently asked by linguists is what the study of the brain could buy them. Clearly, if one is interested in language, one need not be interested in the brain as well. There is absolutely no problem with this position. A problem may arise, however, if one wishes to understand language as a biological system, or as a "brain organ," and still refuses to become concrete about brain processes and representations (see Chomsky, 1980). In this case, it might be advantageous to attempt to connect one's terminology to the putative mechanisms.

It is possible that translation issues can be solved and a language can be developed that refers to linguistic structures and brain structures, linguistic processes and brain processes, and to underlying linguistic principles, as well as to neuroscientific principles of brain structure and function. Given such a language is available, it would be possible to explore the space of possibilities that is restricted on the one side by current neuroscientific knowledge and on the other side by linguistic phenomena. Using neuroscientific knowledge and data for guiding linguistic theorizing appears to be fruitful. Testing syntax theory in neurophysiological experiments may be fruitful as well. Thus, neuroscientific data could then constrain linguistic theory. The reverse may also be true. The study of language may, given such a language connecting noun phrases to neurons (Marshall, 1980) is available, allow for making predictions on brain circuits that have not been detected by other means. Availability of a brain–language interface of this type, a neuronal language theory, may be a necessary condition for deciding between alternative approaches to grammar as it could be a tool for exploring neuron circuits specific to the human brain. A language theory at the neuronal level is required in cognitive neuroscience.

How would the situation, when performing brain imaging studies of language, improve if a brain–language interface were available? It may appear that abstract linguistic theories have some implications for neuroscientific structures and processes. One may therefore claim that interesting questions about brain–language relationships can already be investigated in neuroscientific experiments without additional theoretical work in the first place. The argument here is that this is not true. Linguistic theories do not in fact have strong implications regarding neuroscientific facts, and the experimenter necessarily ends up in difficulties if he or she wishes to entertain them as research guidelines.

A rather fresh view on the relation of neuroscience and linguistics has been put forward by Chomsky (1980) in his book, *Rules and Representations*. Chomsky considered how the presence of an invisible representation of a certain kind during the processing of a sentence could be tested in an electrophysiological experiment, taking as an example the "*wh*- sentence structure." Linguists assign such structures to one type of sentence, *type A*. Now there is a second category of sentences, *type B*, for which such assignment is implied by one of Chomsky's theories, but not by more traditional approaches to syntax. His proposal about the role of a neurophysiological experiment in this theoretical issue is as follows: If "a certain pattern of electrical activity" is highly correlated with the clear cases in which a wh-sentence is processed (type A sentences), and if this same neuronal pattern also occurs during processing of a sentence for which such a representation has been postulated based on theoretical considerations (type B sentence), then one would have evidence that this latter representation of the sentence in question is psychologically and biologically real.

Therefore, in principle, there appears to be no problem with the neuroscientific test of linguistic theories. However, Chomsky's view is not realistic. A closer look at the actual empirical data obtained so far, indicates that clear correlation between language phenomena and patterns of electrical activity are not easy to find. Recent studies of syntactic phenomena have great difficulty in proving that the physiological phenomena that are reported to co-occur with linguistic properties of sentences are strictly related to these linguistic properties per se. As an alternative, they may be related to psychological phenomena in which a linguist is probably less interested, such as working memory load and retrieval (see Kluender & Kutas, 1993). Thus, there would not be a high correlation between a bold brain response and a well-defined linguistic process, but there would be a brain response cooccurring with a variety of linguistic and psychological phenomena.

But this is only the first reason why Chomsky's view is not appropriate. A further important theoretical problem that a physiological test of linguistic ideas must face – ignored by Chomsky – is as follows: The instruments for monitoring brain activity do not by themselves tell the researcher what to look for when investigating linguistic representations and processes.

There are infinite possibilities for describing and analyzing a short time series obtained, for instance, using a multichannel electro- or magnetoencephalograph. What should the language and brain scientist begin with when searching for the pattern that, in the clear case, correlates with the occurrence of a wh- sentence? Answers to this question can be provided only by a theory about the brain mechanisms of language.

A good deal of progress being made in the language and brain sciences is exactly of this nature: finding out what the relevant patterns of brain activity might be. Once again, it must be stressed that brain theory reflections were the seed of such progress.

To mention but one example, it has been speculated by a scholar who further developed the Hebbian brain theory (Milner, 1974) that fast oscillations in the brain may be related to the activity of neuronal ensembles (see also Freeman, 1975; von der Malsburg, 1985). Fast rhythmic brain activity may result from reverberation of neuronal activity in loops of cortical connections that conduct activity in the millisecond range (Chapter 8), or they may emerge from interaction between cell assembly neurons and their inhibitory neighbors (cf. Chapter 2). The idea that high-frequency brain activity is related to the activation of neuronal ensembles and that it may even be a correlate of perceptual and higher cognitive processes inspired numerous experimental investigations in the neurosciences. The results were largely supportive (Bressler & Freeman, 1980; Singer & Gray, 1995). It was therefore proposed that the cell assemblies, or functional webs, representing words produce high-frequency responses when activated. On the basis of this theoretically motivated speculation, several experiments have been conducted, all of which converge on the conclusion that dynamics in high-frequency cortical responses distinguish word forms from meaningless pseudowords (see Chapter 4 for details). More important, the high-frequency responses related to word processes frequently exhibited a specific topography not revealed in other studies of high-frequency responses of cognitive processes (Pulvermüller et al., 1997; Tallon-Baudry & Bertrand, 1999). Furthermore, high-frequency brain responses recorded from motor and visual areas distinguished between words from different categories, and may thus reveal elementary information about more specific properties of word processing in the brain (Pulvermüller, Keil, & Elbert, 1999; Pulvermüller, Lutzenberger et al., 1999).

In the present context, the important conclusion may be that patterns of brain activity possibly indicating aspects of linguistic processing were discovered only because educated guesses about the brain basis of word processing were possible. Without the theorizing these guesses are built on, the probability of finding the patterns would have approximated zero. This is the second reason why Chomsky's idea about a neurobiological test of linguistic constructs is unrealistic: He ignores the biotheoretical prerequisites for such testing.

It can be concluded that although leading linguists may see the possible fruitfulness of neuroscientific research into language, the nature of the major tasks to be addressed in the theoretical neuroscience of language is largely neglected. For serious empirical investigation of the brain mechanisms of language, it is not enough to provide abstract descriptions of language phenomena; it is also necessary to spell out possible language mechanisms in terms of neuronal circuitry.

Building a theory of brain mechanism of language is certainly not an easy task. It is clear that such theorizing will not immediately result in the ultimate answer to the relevant language-related questions. Most likely, early proposals will be falsified by experimental brain research – as, for example, Helmholtz's idea that "Gravitationsenergie" causes the sun to shine had to be replaced. However, the important point at this stage may be to make theoretically relevant brain research on language possible. Scientific investigation of the interesting questions in linguistics requires a brain model of the relevant linguistic processes. The purpose of this book is to give it a try.

References

Abeles, M. (1991). *Corticonics – Neural circuits of the cerebral cortex*. Cambridge: Cambridge University Press.

Abeles, M., Bergman, H., Gat, I., Meilijson, I., Seidemann, E., Tishby, N., & Vaadia, E. (1995). Cortical activity flips among quasi-stationary states. *Proceedings of the National Academy of Sciences, USA, 92*, 8616–20.

Abeles, M., Bergman, H., Margalit, E., & Vaadia, E. (1993). Spatiotemporal firing patterns in the frontal cortex of behaving monkeys. *Journal of Neurophysiology, 70*, 1629–38.

Aboitiz, F., Scheibel, A. B., Fisher, R. S., & Zaidel, E. (1992). Fiber composition of the human corpus callosum. *Brain Research, 598*, 143–53.

Ahissar, E., Vaadia, E., Ahissar, M., Bergman, H., Arieli, A., & Abeles, M. (1992). Dependence of cortical plasticity on correlated activity of single neurons and on behavior context. *Science, 257*, 1412–15.

Ajukiewicz, K. (1936). Die syntaktische Konnexität. *Studia Philosophica, 1*, 1–27.

Angrilli, A., Dobel, C., Rockstroh, B., Stegagno, L., & Elbert, T. (2000). EEG brain mapping of phonological and semantic tasks in Italian and German languages. *Clinical Neurophysiology, 111*, 706–16.

Assadollahi, R., & Pulvermüller, F. (2001). Neuromagnetic evidence for early access to cognitive representations. *Neuroreport, 12*, 207–13.

Bach, E., Brown, C., & Marslen-Wilson, W. (1986). Crossed and nested dependencies in German and Dutch: a psycholinguistic study. *Language and Cognitive Processes, 1*, 249–62.

Bak, T. H., O'Donovan, D. G., Xuereb, J. H., Boniface, S., & Hodges, J. R. (2001). Selective impairment of verb processing associated with pathological changes in Brodmann areas 44 and 45 in the motor neuron disease–dementia–aphasia syndrome. *Brain, 124*, 103–20.

Baker, G. P., & Hacker, P. M. S. (1984). *Language, sense and nonsense*. Oxford: Basil Blackwell.

Bar-Hillel, Y., Perles, M., & Shamir, E. (1961). On formal properties of simple phrase structure grammars. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung, 14*, 143–72.

Barlow, H. (1972). Single units and cognition: a neurone doctrine for perceptual psychology. *Perception, 1*, 371–94.

Barlow, H. B., Hill, R. M., & Levick, W. R. (1964). Retinal ganglion cells responding selectively to direction and speed of image motion in the rabbit. *Journal of Physiology, 173*, 377–407.

Barlow, H., & Levick, W. R. (1965). The mechanism of directionally selective units in rabbit's retina. *Journal of Physiology, 178*, 477–504.

Bavelier, D., Corina, D. P., & Neville, H. J. (1998). Brain and language: a perspective from sign language. *Neuron, 21*, 275–8.

Benson, D. F. (1979). Neurologic correlates of anomia. In H. A. Whitaker & H. Whitaker (Eds.), *Studies in neurolinguistics. Vol. 4*. New York: Academic Press.

Bienenstock, E. (1996). On the dimensionality of cortical graphs. *Journal of Physiology, Paris, 90*(3–4), 251–6.

Bird, H., Lambon-Ralph, M. A., Patterson, K., & Hodges, J. R. (2000). The rise and fall of frequency and imageability: noun and verb production in semantic dementia. *Brain and Language, 73*, 17–49.

Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology, 18*, 355–87.

Bock, J. K., Loebell, H., & Morey, R. (1992). From conceptual roles to structural relations: bridging the syntactic cleft. *Psychological Review, 99*, 150–71.

Braitenberg, V. (1971). *On the texture of brains*. Heidelberg: Springer.

Braitenberg, V. (1978a). Cell assemblies in the cerebral cortex. In R. Heim & G. Palm (Eds.), *Theoretical approaches to complex systems. (Lecture notes in biomathematics, vol. 21)* (pp. 171–88). Berlin: Springer.

Braitenberg, V. (1978b). Cortical architectonics: general and areal. In M. A. B. Brazier & H. Petsche (Eds.), *Architectonics of the cerebral cortex* (pp. 443–65). New York: Raven Press.

Braitenberg, V. (1980). Alcune considerazione sui meccanismi cerebrali del linguaggio. In G. Braga & V. Braitenberg & C. Cipolli & E. Coseriu & S. Crespi-Reghizzi & J. Mehler & R. Titone (Eds.), *L'accostamento interdisciplinare allo studio del linguaggio* (pp. 96–108). Milano: Franco Angeli Editore.

Braitenberg, V. (1996). *Il gusto della lingua. Meccanismi cerebrali e strutture grammaticali*. Merano: Alpha & Beta.

Braitenberg, V., Heck, D., & Sultan, F. (1997). The detection and generation of sequences as a key to cerebellar function: experiments and theory. *Behavioral and Brain Sciences, 20*, 229–45.

Braitenberg, V., & Pulvermüller, F. (1992). Entwurf einer neurologischen Theorie der Sprache. *Naturwissenschaften, 79*, 103–17.

Braitenberg, V., & Schüz, A. (1998). *Cortex: statistics and geometry of neuronal connectivity* (2 ed.). Berlin: Springer.

Brent, M. R., & Cartwright, T. A. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition, 61*(1–2), 93–125.

Bressler, S. L. (1995). Large-scale cortical networks and cognition. *Brain Research Review, 20*, 288–304.

Bressler, S. L., & Freeman, W. J. (1980). Frequency analysis of olfactory system EEG in cat, rabbit and rat. *Electroencephalography and Clinical Neurophysiology, 50*, 19–24.

Broca, P. (1861). Remarques sur la siège de la faculté de la parole articulée, suivies d'une observation d'aphémie (perte de parole). *Bulletin de la Société d'Anatomie, 36*, 330–57.

Brodmann, K. (1909). *Vergleichende Lokalisationslehre der Großhirnrinde*. Leipzig: Barth.

Brown, C. M., Hagoort, P., & ter Keurs, M. (1999). Electrophysiological signatures of visual lexical processing: open- and closed-class words. *Journal of Cognitive Neuroscience, 11*, 261–81.

Brown, T. H., Kairiss, E. W., & Keeman, C. L. (1996). Hebbian synapses: biophysical mechanisms and algorithms. *Review in Neuroscience, 13*, 475–511.

Brown, W. S., & Lehmann, D. (1979). Verb and noun meaning of homophone words activate different cortical generators: a topographic study of evoked potential fields. *Experimental Brain Research, 2*, S159–68.

Bryden, M. P., Hecaen, H., & DeAgostini, M. (1983). Patterns of cerebral organization. *Brain and Language, 20*(2), 249–62.

Buonomano, D. V. (2000). Decoding temporal information: A model based on short-term synaptic plasticity. *J Neuroscience, 20*, 1129–41.

Buonomano, D. V., & Merzenich, M. M. (1998). Cortical plasticity: from synapses to maps. *Annual Review of Neuroscience, 21*, 149–86.

Caplan, D. (1987). *Neurolinguistics and linguistic aphasiology. An introduction*. Cambridge, MA: Cambridge University Press.

Caplan, D. (1996). *Language: structure, processing, and disorders*. Cambridge, MA: MIT Press.

Cappa, S. F., Binetti, G., Pezzini, A., Padovani, A., Rozzini, L., & Trabucchi, M. (1998). Object and action naming in Alzheimer's disease and frontotemporal dementia. *Neurology, 50*(2), 351–5.

Caramazza, A., & Zurif, E. B. (1976). Dissociation of algorithmic and heuristic processes in sentence comprehension: evidence from aphasia. *Brain and Language, 3*, 572–82.

Charniak, E. (1993). *Statistical language learning*. Cambridge, MA: MIT Press.

Cheour, M., Ceponiene, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K., & Näätänen, R. (1998). Development of language-specific phoneme representations in the infant brain. *Nature Neuroscience, 1*, 351–3.

Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.

Chomsky, N. (1963). Formal properties of grammars. In R. D. Luce & R. R. Bush & E. Galanter (Eds.), *Handbook of mathematical psychology, Vol. 2* (pp. 323–418). New York, London: Wiley.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Chomsky, N. (1980). *Rules and representations*. New York: Columbia University Press.

Chomsky, N. (2000). *New horizons in the study of language and mind*. Cambridge, MA: Cambridge University Press.

Creutzfeldt, O., Ojemann, G., & Lettich, E. (1989). Neuronal activity in the human lateral temporal lobe. I. Responses to speech. *Experimental Brain Research, 77*, 451–75.

Cruse, H., Bartling, C., Cymbalyuk, G., Dean, J., & Dreifert, M. (1995). A modular artificial neural net for controlling a six-legged walking system. *Biological Cybernetics, 72*, 421–30.

Cruse, H., & Bruwer, M. (1987). The human arm as a redundant manipulator: the control of path and joint angles. *Biological Cybernetics, 57*, 137–44.

Dale, A. M., Liu, A. K., Fischl, B. R., Buckner, R. L., Belliveau, J. W., Lewine, J. D., & Halgren, E. (2000). Dynamic statistical parametric mapping: combining fMRI and MEG for high-resolution imaging of cortical activity. *Neuron, 26*(1), 55–67.

Damasio, A. R. (1989). The brain binds entities and events by multiregional activation from convergence zones. *Neural Computation, 1*, 123–32.

Damasio, H., Grabowski, T. J., Tranel, D., Hichwa, R. D., & Damasio, A. R. (1996). A neural basis for lexical retrieval. *Nature, 380*, 499–505.

Damasio, A. R., & Tranel, D. (1993). Nouns and verbs are retrieved with differently distributed neural systems. *Proceedings of the National Academy of Sciences, USA, 90*, 4957–60.

Daniele, A., Giustolisi, L., Silveri, M. C., Colosimo, C., & Gainotti, G. (1994). Evidence for a possible neuroanatomical basis for lexical processing of nouns and verbs. *Neuropsychologia, 32*, 1325–41.

Deacon, T. W. (1992). Cortical connections of the inferior arcuate sulcus cortex in the macaque brain. *Brain Research, 573*(1), 8–26.

DeFelipe, J., & Farinas, I. (1992). The pyramidal neuron of the cerebral cortex: morphological and chemical characteristics of the synaptic inputs. *Progress in Neurobiology, 39*(6), 563–607.

Dehaene, S., Changeux, J. P., & Nadal, J. P. (1987). Neural networks that learn temporal sequences by selection. *Proceedings of the National Academy of Sciences, USA, 84*, 2727–31.

Dehaene-Lambertz, G., & Dehaene, S. (1994). Speed and cerebral correlates of syllable discrimination in infants. *Nature, 370*, 292–5.

Dell, G. S. (1986). A spreading-activation theory of retreival in sentence production. *Psychological Review, 93*, 283–321.

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review, 104*(4), 801–38.

De Renzi, E., & Vignolo, L. (1962). The Token Test: a sensitive test to detect receptive disturbances in aphasics. *Brain, 85*, 665–78.

de Saussure, F. (1916). *Cours de linguistique generale*. Paris: Payot.

Devlin, J. T., Russell, R. P., Davis, M. H., Price, C. J., Moss, H. E., Fadili, M. J., & Tyler, L. K. (2002). Is there an anatomical basis for category-specificity? Semantic memory studies in PET and fMRI. *Neuropsychologia, 40*, 54–75.

Diesch, E., Biermann, S., & Luce, T. (1998). The magnetic mismatch field elicited by words and phonological non-words. *NeuroReport, 9*, 455–60.

Diesmann, M., Gewaltig, M. O., & Aertsen, A. (1999). Stable propagation of synchronous spiking in cortical neural networks. *Nature, 402*(6761), 529–33.

Dilthey, W. (1989). *Einleitung in die Geisteswissenschaften [Introduction to the human sciences], Vol. 1*. Princeton, NJ: Princeton University Press.

Dobel, C., Pulvermüller, F., Härle, M., Cohen, R., Kobbel, P., Schönle, P. W., & Rockstroh, B. (2001). Syntactic and semantic processing in the healthy and aphasic human brain. *Experimental Brain Research, 140*, 77–85.

Egelhaaf, M., Borst, A., & Reichardt, W. (1989). Computational structure of a biological motion-detection system as revealed by local detector analysis in the fly's nervous system. *Journal of the Optical Society of America (A), 6*, 1070–87.

Eisenberg, P. (1999). *Grundriss der deutschen Grammatik: Der Satz*, Vol. 2. Stuttgart: Verlag J.B. Metzler.

Elbert, T., Pantev, C., Wienbruch, C., Rockstroh, B., & Taub, E. (1995). Increased cortical representation of the fingers of the left hand in string players. *Science, 270*(5234), 305–7.

Ellis, A. W., & Young, A. W. (1988). *Human cognitive neuropsychology*. Hove, UK: Lawrence Erlbaum Associates Ltd.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science, 14*, 179–211.

Elman, J. L., Bates, L., Johnson, M., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: a connectionist perspective on development*. Cambridge, MA: MIT Press.

Engert, F., & Bonhoeffer, T. (1999). Dendritic spine changes associated with hippocampal long-term synaptic plasticity. *Nature, 399*(6731), 66–70.

Epstein, H. T. (1999). Other brain effects of words. *Behavioral and Brain Sciences, 22*, 287–8.

Eulitz, C., Eulitz, H., Maess, B., Cohen, R., Pantev, C., & Elbert, T. (2000). Magnetic brain activity evoked and induced by visually presented words and nonverbal stimuli. *Psychophysiology, 37*(4), 447–55.

Farah, M. J., & McClelland, J. L. (1991). A computational model of semantic memory impairment: modality specificity and emergent category specificity. *Journal of Experimental Psychology: General, 120*, 339–57.

Federmeier, K. D., Segal, J. B., Lombrozo, T., & Kutas, M. (2000). Brain responses to nouns, verbs and class-ambiguous words in context. *Brain, 123*, 2552–66.

Fiez, J. A., & Petersen, S. E. (1998). Neuroimaging studies of word reading. *Proceedings of the National Academy of Science, USA, 95*(3), 914–21.

Fiez, J. A., Raichle, M. E., Balota, D. A., Tallal, P., & Petersen, S. E. (1996). PET activation of posterior temporal regions during auditory word presentation and verb generation. *Cerebral Cortex, 6*, 1–10.

Freeman, W. J. (1975). *Mass action in the nervous system*. New York: Academic Press.

Freud, S. (1891). *Zur Auffassung der Aphasien*. Leipzig, Wien: Franz Deuticke.

Friederici, A., & Mecklinger, A. (1996). Syntactic parsing as revealed by brain responses: first pass and second pass parsing processes. *Journal of Psycholinguistic Research, 25*, 157–76.

Friederici, A. D. (1997). Neurophysiological aspects of language processing. *Clinical Neuroscience, 4*, 64–72.

Friederici, A. D., Pfeifer, E., & Hahne, A. (1993). Event-related brain potentials during natural speech processing: effects of semantic, morphological and syntactic violations. *Cognitive Brain Research, 1*, 183–92.

Fry, D. B. (1966). The development of the phonological system in the normal and deaf child. In F. Smith & G. A. Miller (Eds.), *The genesis of language* (pp. 187–206). Cambridge, MA: MIT Press.

Fuster, J. M. (1995). *Memory in the cerebral cortex. An empirical approach to neural networks in the human and nonhuman primate*. Cambridge, MA: MIT Press.

Fuster, J. M. (1997). Network memory. *Trends in Neurosciences, 20*, 451–9.

Fuster, J. M. (1998a). Distributed memory for both short and long term. *Neurobiology in Learning and Memory, 70*(1–2), 268–74.

Fuster, J. M. (1998b). Linkage at the top. *Neuron, 21*(6), 1223–4.

Fuster, J. M. (1999). Hebb's other postulate at work on words. *Behavioral and Brain Sciences, 22*, 288–9.

Fuster, J. M. (2000). Cortical dynamics of memory. *International Journal of Psychophysiology, 35*(2–3), 155–64.

Fuster, J. M., & Alexander, G. E. (1971). Neuron activity related to short-term memory. *Science, 173*(997), 652–4.

Fuster, J. M., Bodner, M., & Kroger, J. K. (2000). Cross-modal and cross-temporal association in neurons of frontal cortex. *Nature, 405*(6784), 347–51.

Fuster, J. M., & Jervey, J. P. (1982). Neuronal firing in the inferiotemporal cortex of the monkey in a visual memory task. *Journal of Neuroscience, 2*, 361–75.

Gaifman, C. (1965). Dependency systems and phrase structure systems. *Information and Control, 8*, 304–37.

Galuske, R. A., Schlote, W., Bratzke, H., & Singer, W. (2000). Interhemispheric asymmetries of the modular structure in human temporal cortex. *Science, 289*(5486), 1946–9.

Garrett, M. (1980). Levels of processing in sentence production. In B. Butterworth (Ed.), *Language Production I* (pp. 177–220). London: Academic Press.

Gazdar, G., Klein, E., Pullum, G., & Sag, I. (1985). *Generalized phrase structure grammar*. Cambridge, MA: Harvard University Press.

Geschwind, N. (1965a). Disconnection syndromes in animals and man (1). *Brain, 88*, 237–94.

Geschwind, N. (1965b). Disconnection syndromes in animals and man (2). *Brain, 88*, 585–644.

Geschwind, N. (1974). *Selected papers on language and the brain*. Dordrecht: Kluewer.

Geschwind, N., & Levitsky, W. (1968). Human brain: left–right asymmetries in temporal speech region. *Science, 161*, 186–7.

Glaser, W. R., & Düngelhoff, F. J. (1984). The time course of picture–word interference. *Journal of Experimental Psychology: Human Perception and Performance, 10*, 640–54.

Goodglass, H., & Quadfasel, F. A. (1954). Language laterality in left-handed aphasics. *Brain, 77*, 521–48.

Grabowski, T. J., Damasio, H., & Damasio, A. R. (1998). Premotor and prefrontal correlates of category-related lexical retrieval. *Neuroimage, 7*, 232–43.

Gunter, T. C., Friederici, A. D., & Schriefers, H. (2000). Syntactic gender and semantic expectancy: ERPs reveal early autonomy and late interaction. *Journal of Cognitive Neuroscience, 12*(4), 556–68.

Haegeman, L. (1991). *Introduction to government and binding theory*. Cambridge, MA: Basil Blackwell.

Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP-measure of syntactic processing. *Language and Cognitive Processes, 8*, 439–483.

Hagoort, P., Indefrey, P., Brown, C., Herzog, H., Steinmetz, H., & Seitz, R. J. (1999). The neural circuitry involved in the reading of German words and pseudowords: a PET study. *Journal of Cognitive Neuroscience, 11*(4), 383–98.

Hahne, A., & Friederici, A. D. (1999). Electrophysiological evidence for two steps in syntactic analysis. Early automatic and late controlled processes. *Journal of Cognitive Neuroscience, 11*(2), 194–205.

Hare, M., Elman, J. L., & Daugherty, K. G. (1995). Default generalisation in connectionist networks. *Language and Cognitive Processes, 10*, 601–30.

Harris, Z. S. (1945). Discontinuous morphemes. *Language, 21*, 121–7.

Harris, Z. S. (1951). *Structural linguistics*. Chicago: Chicago University Press.

Harris, Z. S. (1952). Discourse analysis. *Language, 28*, 1–30.

Harris, Z. S. (1955). From phonemes to morphemes. *Language, 31*, 190–222.

Harris, Z. S. (1957). Co-occurrence and transformation in linguistic structure. *Language, 33*, 283–340.

Harris, Z. S. (1965). Transformational theory. *Language, 41*, 363–401.

Hasbrooke, R. E., & Chiarello, C. (1998). Bihemispheric processing of redundant bilateral lexical information. *Neuropsychology, 12*, 78–94.

Hauk, O., & Pulvermüller, F. (2002). Neurophysiological distinction of action words in the frontal lobe: an ERP study using minimum current estimates.

Hayes, T. L., & Lewis, D. (1993). Hemispheric differences in layer III pyramidal neurons of the anterior language area. *Archives of Neurology, 50*, 501–5.

Hays, D. G. (1964). Dependency theory: a formalism and some observations. *Language, 40*, 511–25.

Hebb, D. O. (1949). *The organization of behavior. A neuropsychological theory*. New York: John Wiley.

Hecaen, H., De Agostini, M., & Monzon-Montes, A. (1981). Cerebral organization in left-handers. *Brain and Language, 12*(2), 261–84.

Hellwig, B. (2000). A quantitative analysis of the local connectivity between pyramidal neurons in layers 2/3 of the rat visual cortex. *Biological Cybernetics, 82*(2), 111–21.

Heringer, H. J. (1996). *Deutsche Syntax dependentiell*. Tübingen: Stauffenburg Verlag.

Hetherington, P. A., & Shapiro, M. L. (1993). Simulating Hebb cell assemblies: the necessity for partitioned dendritic trees and a post-not-pre LTD rule. *Network, 4*, 135–53.

Hinton, G. E., & Shallice, T. (1991). Lesioning an attractor network: investigation of acquired dyslexia. *Psychological Review, 98*, 74–95.

Hubel, D. (1995). *Eye, brain, and vision* (2 ed.). New York: Scientific American Library.

Humphreys, G. W., Evett, L. J., & Taylor, D. E. (1982). Automatic phonological priming in visual word recognition. *Memory and Cognition, 10*, 576–90.

Humphreys, G. W., & Forde, E. M. E. (2001). Hierarchies, similarity and interactivity in object recognition. *Behavioral and Brain Sciences, 24*, 453–509

Indefrey, P., Hagoort, P., Herzog, H., Seitz, R. J., & Brown, C. M. (2001). Syntactic processing in left prefrontal cortex is independent of lexical meaning. *Neuroimage, 14*, 546–55.

Jackendoff, R. (1977). *X-bar syntax: a study of phrase structure*. Cambridge, MA: MIT Press.

Jackson, J. H. (1878). On affections of speech from disease of the brain (1). *Brain, 1*, 304–30.

Jacobs, B., Batal, H. A., Lynch, B., Ojemann, G., Ojemann, L. M., & Scheibel, A. B. (1993). Quantitative dendritic and spine analyses of speech cortices: a case study. *Brain and Language, 44*, 239–53.

Joanette, Y., Goulet, P., & Hannequin, D. (1990). *Right hemisphere and verbal communication*. New York: Springer-Verlag.

Joshi, A. K., & Levy, L. S. (1982). Phrase structure trees bear more fruit than you would have thought. *American Journal of Computational Linguistics, 8*, 1–11.

Kandel, E. R. (1991). Cellular mechanisms of learning and the biological basis of individuality. In E. R. Kandel & J. H. Schwartz & T. M. Jessell (Eds.), *Principles of neural science* (3 ed., pp. 1009–31). New York: Elsevier.

Kaplan, R. M. (1972). Augmented transition networks as psychological models of sentence comprehension. *Artificial Intelligence, 3*, 77–100.

Kiefer, M. (2001). Perceptual and semantic sources of category-specific effects: Event-related potentials during picture and word categorization. *Memory and Cognition, 29*, 100–16.

Kiefer, M., & Spitzer, M. (2001). The limits of a distributed account of conceptual knowledge. *Trends in Cognitive Sciences, 5*, 469–71.

Kleene, S. C. (1956). Representation of events in nerve nets and finite automata. In C. E. Shannon & J. McCarthy (Eds.), *Automata studies* (pp. 3–41). Princeton, NJ: Princeton University Press.

Kleinfeld, D., & Sompolinsky, H. (1988). Associative neural network model for the generation of temporal patterns. Theory and application to central pattern generators. *Biophysical Journal, 54*, 1039–51.

Kluender, R., & Kutas, M. (1993). Bridging the gap: evidence from ERPs on the processing of unbounded dependencies. *Journal of Cognitive Neuroscience, 5*, 196–214.

Koenig, T., & Lehmann, D. (1996). Microstates in language-related brain potential maps show noun–verb differences. *Brain and Language, 53*, 169–82.

Kolk, H. H. J., van Grunsven, M. J. F., & Keyser, A. (1985). On the parallelism between production and comprehension in agrammatism. In M.-L. Kean (Ed.), *Agrammatism* (pp. 165–206). New York: Academic Press.

Korpilahti, P., Krause, C. M., Holopainen, I., & Lang, A. H. (2001). Early and late mismatch negativity elicited by words and speech-like stimuli in children. *Brain and Language, 76*, 332–39.

Krause, C. M., Korpilahti, P., Porn, B., Jantti, J., & Lang, H. A. (1998). Automatic auditory word perception as measured by 40 Hz EEG responses. *Electroencephalography and Clinical Neurophysiology, 107*, 84–7.

Kreiter, A. K. (2002). Functional implications of temporal structure in primate cortical information processing. *Zoology: Analysis of Complex Systems, 104*, 241–55.

Kujala, T., Alho, K., & Naatanen, R. (2000). Cross-modal reorganization of human cortical functions. *Trends in Neuroscience, 23*(3), 115–20.

Lambek, J. (1958). The mathematics of sentence structure. *Americal Mathematical Monthly, 65*, 154–70.

Lambek, J. (1959). Contributions to a mathematical analysis of the English verb phrase. *Journal of the Canadian Linguistic Association, 5*, 83–9.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: the Latent Semantic Analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104*, 211–40.

Lashley, K. S. (1950). In search of the engram. *Symposium of the Society for Experimental Biology, 4*, 454–82.

Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior. The Hixxon symposium*, pp. 112–36. New York: John Wiley.

Le Clec'H, G., Dehaene, S., Cohen, L., Mehler, J., Dupoux, E., Poline, J. B., Lehericy, S., van de Moortele, P. F., & Le Bihan, D. (2000). Distinct cortical areas for names of numbers and body parts independent of language and input modality. *Neuroimage, 12*, 381–91.

Lee, K. H., Chung, K., Chung, J. M., & Coggeshall, R. E. (1986). Correlation of cell body size, axon size, and signal conduction velocity for individually labelled dorsal root ganglion cells in the cat. *Journal of Comparative Neurology, 243*(3), 335–46.

Lehmann, D., & Skrandies, W. (1984). Spatial analysis of evoked potentials in man: a review. *Progress in Neurobiology, 23*, 227–50.

Levelt, W. J. M. (1974). *Formal grammars in linguistics and psycholinguistics. Vol. 1. An introduction to the theory of formal languages and automata*. The Hague, Paris: Mouton.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*, 1–75.

Levelt, W. J., Schriefers, H., Vorberg, D., Meyer, A. S., Pechmann, T., & Havinga, J. (1991). The time course of lexical access in speech production: A study of picture naming. *Psychological Review, 98*, 122–42.

Lichtheim, L. (1885). On aphasia. *Brain, 7*, 433–84.

Locke, J. L. (1989). Babbling and early speech: continuity and individual differences. *First Language, 9*, 191–206.

Locke, J. L. (1993). *The child's path to spoken language*. Cambridge, MA: Harvard University Press.

Lutzenberger, W., Pulvermüller, F., & Birbaumer, N. (1994). Words and pseudowords elicit distinct patterns of 30-Hz activity in humans. *Neuroscience Letters, 176*, 115–18.

Marcel, A. J., & Patterson, K. (1978). Word recognition and production: reciprocity in clinical and normal studies. In J. Requin (Ed.), *Attention and performance* (Vol. VII, pp. 209–26). New Jersey: Lawrence Erlbaum Associates.

Marcus, S. (1965). Sur la Notion de la Projectivite. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik, 11*, 181–92.

Marie, P. (1906). Revision de la question de l'aphasie: la triosieme circonvolution frontale gauche ne joue uncun role special dans la fonction du langage. *Semaine Medicale (Paris), 26*, 241–7.

Markov, A. A. (1913). Essai d'une recherche statistique sur le texte du roman "Eugene Onegin." *Bulletin de l'Academie Imperiale des Sciences, St. Petersburg, 7*.

Marshall, J. C. (1980). On the biology of language acquisition. In D. Caplan (Ed.), *Biological studies of mental processes*. Cambridge, MA: MIT Press.

Marslen-Wilson, W., & Tyler, L. (1997). Dissociating types of mental computation. *Nature, 387*, 592–4.

Marslen-Wilson, W., & Tyler, L. K. (1975). Processing structure of sentence perception. *Nature, 257*(5529), 784–6.

Marslen-Wilson, W., Tyler, L. K., Waksler, R., & Older, L. (1994). Morphology and meaning in the English mental lexicon. *Psychological Review, 101*, 3–33.

Marslen-Wilson, W., & Warren, P. (1994). Levels of perceptual representation and process in lexical access: words, phonemes, and features. *Psychological Review, 101*, 633–75.

Marslen-Wilson, W. D., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition, 8*, 1–71.

Martin, A., & Chao, L. L. (2001). Semantic memory and the brain: structure and processes. *Current Opinion in Neurobiology, 11*, 194–201.

Martin, A., Wiggs, C. L., Ungerleider, L. G., & Haxby, J. V. (1996). Neural correlates of category-specific knowledge. *Nature, 379*, 649–52.

Mason, A., Nicoll, A., & Stratford, K. (1991). Synaptic transmission between individual pyramidal neurons of the rat visual cortex in vitro. *Journal of Neuroscience, 11*(1), 72–84.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review, 88*, 375–407.

McCulloch, W. S., & Pitts, W. H. (1943). A logical calculus of ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics, 5*, 115–33.

Merzenich, M. M., Kaas, J. H., Wall, J., Nelson, R. J., Sur, M., & Felleman, D. (1983). Topographic reorganization of somatosensory cortical areas 3b and 1 in adult monkeys following restricted deafferentation. *Neuroscience, 8*(1), 33–55.

Merzenich, M. M., Kaas, J. H., Wall, J. T., Sur, M., Nelson, R. J., & Felleman, D. J. (1983). Progression of change following median nerve section in the cortical representation of the hand in areas 3b and 1 in adult owl and squirrel monkeys. *Neuroscience, 10*(3), 639–65.

Mesulam, M. M. (1990). Large-scale neurocognitive networks and distributed processing for attention, language, and memory. *Annals of Neurology, 28*, 597–613.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: eviedence of a dependence of retrieval operations. *Journal of Experimental Psychology, 90*, 227–35.

Miceli, G., Silveri, C., Nocentini, U., & Caramazza, A. (1988). Patterns of dissociation in comprehension and production of nouns and verbs. *Aphasiology, 2*, 351–8.

Miceli, G., Silveri, M., Villa, G., & Caramazza, A. (1984). On the basis of agrammatics' difficulty in producing main verbs. *Cortex, 20*, 207–20.

Miller, G. A., & Chomsky, N. (1963). Finite models of language users. In R. D. Luce & R. R. Bush & E. Galanter (Eds.), *Handbook of mathematical psychology, Vol. 2* (pp. 419–91). New York, London: Wiley.

Miller, R. (1991). *Cortico-hippocampal interplay and the representation of contexts in the brain*. Berlin: Springer.

Miller, R. (1996). *Axonal conduction times and human cerebral laterality. A psychobiological theory*. Amsterdam: Harwood Academic Publishers.

Miller, R., & Wickens, J. R. (1991). Corticostriatal cell assemblies in selective attention and in representation of predictable and controllable events: a general statement of corticostriatal interplay and the role of striatal dopamine. *Concepts in Neuroscience, 2*, 65–95.

Milner, P. M. (1957). The cell assembly: Mk. II. *Psychological Review, 64*, 242–52.

Milner, P. M. (1974). A model for visual shape recognition. *Psychological Review, 81*, 521–35.

Milner, P. M. (1996). Neural representation: some old problems revisited. *Journal of Cognitive Neuroscience, 8*, 69–77.

Minsky, M. (1972). *Computation: finite and infinite machines*. London: Prentice-Hall.

Minsky, M., & Papert, S. (1969). *Perceptrons*. Cambridge, MA: MIT Press.

Mohr, B., & Pulvermüller, E. (2002). Redundancy gains and costs in word processing: the effect of short stimulus onset asynchronies (SOAs). *Journal of Experimental Psychology: Learning, Memory and Cognition.*

Mohr, B., Pulvermüller, F., Mittelstädt, K., & Rayman, J. (1996). Multiple stimulus presentation facilitates lexical processing. *Neuropsychologia, 34*, 1003–13.

Mohr, B., Pulvermüller, F., Rayman, J., & Zaidel, E. (1994). Interhemispheric cooperation during lexical processing is mediated by the corpus callosum: evidence from the split-brain. *Neuroscience Letters, 181*, 17–21.

Mohr, B., Pulvermüller, F., & Zaidel, E. (1994). Lexical decision after left, right and bilateral presentation of content words, function words and non-words: evidence for interhemispheric interaction. *Neuropsychologia, 32*, 105–24.

Molfese, D. L. (1972). *Cerebral asymmetry in infants, children and adults: auditory evoked responses to speech and noise stimuli.* Unpublished doctoral dissertation: Pennsylvania State University.

Molfese, D. L., Burger-Judisch, L. M., Gill, L. A., Golinkoff, R. M., & Hirsch-Pasek, K. A. (1996). Electrophysiological correlates of noun-verb processing in adults. *Brain and Language, 54*, 388–413.

Montoya, P., Larbig, W., Pulvermüller, F., Flor, H., & Birbaumer, N. (1996). Cortical correlates of semantic classical conditioning. *Psychophysiology, 33*, 644–49.

Moore, C. J., & Price, C. J. (1999). A functional neuroimaging study of the variables that generate category-specific object processing differences. *Brain, 122*, 943–62.

Moro, A., Tettamanti, M., Perani, D., Donati, C., Cappa, S. F., & Fazio, F. (2001). Syntax and the brain: disentangling grammar by selective anomalies. *Neuroimage, 13*, 110–18.

Morton, J. (1969). The interaction of information in word recognition. *Psychological Review, 76*, 165–78.

Müller, H. M., King, J. W., & Kutas, M. (1997). Event-related potentials elicited by spoken relative clauses. *Cognitive Brain Research, 5*, 193–203.

Mummery, C. J., Patterson, K., Hodges, J. R., & Price, C. J. (1998). Functional neuroanatomy of the semantic system: divisible by what? *Journal of Cognitive Neuroscience, 10*, 766–77.

Näätänen, R. (2001). The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology, 38*, 1–21.

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., Valnio, A., Alku, P., Ilmoniemi, R. J., Luuk, A., Allik, J., Sinkkonen, J., & Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature, 385*, 432–4.

Näätänen, R., & Winkler, I. (1999). The concept of auditory stimulus representation in cognitive neuroscience. *Psychological Bulletin, 12*, 826–59.

Neininger, B., & Pulvermüller, F. (2001). The right hemisphere's role in action word processing: a double case study. *Neurocase, 7*, 303–17.

Neininger, B., & Pulvermüller, E. (2002). Word category–specific deficits after lesions in the right hemisphere. *Neuropsychologia,* in press.

Neville, H., Nicol, J. L., Barss, A., Forster, K. I., & Garrett, M. F. (1991). Syntactically based sentence processing classes: evidence from event-related brain potentials. *Journal of Cognitive Neuroscience, 3*, 151–65.

Neville, H. J., Bavelier, D., Corina, D., Rauschecker, J., Karni, A., Lalwani, A., Braun, A., Clark, V., Jezzard, P., & Turner, R. (1998). Cerebral organization for language in deaf and hearing subjects: biological constraints and effects of experience. *Proceedings of the National Academy of Sciences, USA, 95*, 922–9.

Neville, H. J., Mills, D. L., & Lawson, D. S. (1992). Fractionating language: different neural subsystems with different sensitive periods. *Cerebral Cortex, 2*, 244–58.

Newman, A. J., Bavelier, D., Corina, D., Jezzard, P., & Neville, H. J. (2002). A critical period for right hemisphere recruitment in American Sign Language processing. *Neuroscience, 5*, 76–80.

Nobre, A. C., & McCarthy, G. (1994). Language-related EPRs: scalp distributions and modulation by word type and semantic priming. *Journal of Cognitive Neuroscience, 6*, 233–255.

Noppeney, U., & Price, C. J. (2002). Retrieval of visual, auditory, and abstract semantics. *Neuroimage, 15*, 917–26.

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: feedback is never necessary. *Behavioral and Brain Sciences, 23*, 299–370.

Osterhout, L. (1997). On the brain response to syntactic anomalies: manipulations of word position and word class reveal individual differences. *Brain and Language, 59*(3), 494–522.

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language, 31*, 785–806.

Page, M. P., & Norris, D. (1998). The primacy model: a new model of immediate serial recall. *Psychological Review, 105*, 761–81.

Palm, G. (1980). On associative memory. *Biology and Cybernetics, 36*(1), 19–31.

Palm, G. (1981). Towards a theory of cell assemblies. *Biology and Cybernetics, 39*(3), 181–94.

Palm, G. (1982). *Neural assemblies*. Berlin: Springer.

Palm, G. (1990). Local learning rules and sparse coding in neural networks. In R. Eckmiller (Ed.), *Advanced neural computers* (pp. 145–50). Amsterdam: Elsevier.

Palm, G. (1993a). Cell assemblies, coherence, and corticohippocampal interplay. *Hippocampus, 3*(Spec No), 219–25.

Palm, G. (1993b). On the internal structure of cell assemblies. In A. Aertsen (Ed.), *Brain theory: spatio-temporal aspects of brain function* (pp. 261–70). Amsterdam: Elsevier.

Palm, G., & Sommer, F. T. (1995). Associative data storage and retrieval in neural networks. In E. Domany & J. L. van Hemmen & K. Schulten (Eds.), *Models of neural networks III* (pp. 79–118). New York: Springer Verlag.

Pandya, D. N., & Yeterian, E. H. (1985). Architecture and connections of cortical association areas. In A. Peters & E. G. Jones (Eds.), *Cerebral cortex, Vol. 4. Association and auditory cortices* (pp. 3–61). London: Plenum Press.

Patterson, K., & Hodges, J. R. (2001). Semantic dementia. In J. L. McClelland (Ed.), *The international encyclopaedia of the social and behavioural sciences. Section: disorders of the adult brain.* New York: Pergamon Press.

Paus, T., Perry, D. W., Zatorre, R. J., Worsley, K. J., & Evans, A. C. (1996). Modulation of cerebral blood flow in the human auditory cortex during speech: role of motor-to-sensory discharges. *European Journal of Neuroscience, 8*, 2236–46.

Penfield, W., & Rassmussen, T. (1950). *The cerebral cortex of man*. New York: Macmillan.

Penfield, W., & Roberts, L. (1959). *Speech and brain mechanisms*. Princeton, NJ: Princeton University Press.

Perani, D., Schnur, T., Tettamanti, M., Gorno-Tempini, M., Cappa, S. F., & Fazio, F. (1999). Word and picture matching: a PET study of semantic category effects. *Neuropsychologia, 37*, 293–306.

Petersen, S., & Fiez, J. A. (1993). The processing of single words studied with positron emission tomography. *Annual Review in Neuroscience, 16*, 509–30.

Petersen, S., Fox, P., Posner, M., Mintun, M., & Raichle, M. (1989). Positron emission tomography studies of the processing of single words. *Journal of Cognitive Neuroscience, 1*, 153–70.

Petitto, L. A., Zatorre, R. J., Gauna, K., Nikelski, E. J., Dostie, D., & Evans, A. C. (2000). Speech-like cerebral activity in profoundly deaf people processing signed languages: implications for the neural basis of human language. *Proceedings of the National Academy of Sciences, USA, 97*, 13961–6.

Petri, C. A. (1970). *Kommunikation mit Automaten*. Dissertation: Universität Bonn.

Pick, A. (1913). *Die agrammatischen Sprachstörungen. Studien zur psychologischen Grundlegung der Aphasielehre*. Berlin: Springer.

Pickering, M. J., & Branigan, H. P. (1999). Syntactic priming in language production. *Trends in Cognitive Sciences, 3*, 136–41.

Pinker, S. (1994). *The language instinct. How the mind creates language*. New York: Harper Collins Publishers.

Pinker, S. (1997). Words and rules in the human brain. *Nature, 387*, 547–8.

Plaut, D. C., & Shallice, T. (1993). Deep dyslexia: a case study of connectionist neuropsychology. *Cognitive Neuropsychology, 10*, 377–500.

Poizner, H., Bellugi, U., & Klima, E. S. (1990). Biological foundations of language: Clues from sign language. *Annual Review in Neuroscience, 13*, 283–307.

Posner, M. I., & DiGirolamo, G. J. (1999). Flexible neural circuitry in word processing. *Behavioral and Brain Sciences, 22*, 299–300.

Preissl, H., Pulvermüller, F., Lutzenberger, W., & Birbaumer, N. (1995). Evoked potentials distinguish nouns from verbs. *Neuroscience Letters, 197*, 81–3.

Previc, F. H. (1991). A general theory concerning the prenatal origins of cerebral lateralization in humans. *Psychological Review, 98*, 299–334.

Price, C. J. (2000). The anatomy of language: contributions from functional neuroimaging. *Journal of Anatomy, 197 Pt 3*, 335–59.

Price, C. J., Warburton, E. A., Moore, C. J., Frackowiak, R. S., & Friston, K. J. (2001). Dynamic diaschisis: anatomically remote and context-sensitive human brain lesions. *Journal of Cognitive Neuroscience, 13*, 419–29.

Price, C. J., Wise, R. J. S., & Frackowiak, R. S. J. (1996). Demonstrating the implicit processing of visually presented words and pseudowords. *Cerebral Cortex, 6*, 62–70.

Pulvermüller, F. (1992). Constituents of a neurological theory of language. *Concepts in Neuroscience, 3*, 157–200.

Pulvermüller, F. (1993). On connecting syntax and the brain. In A. Aertsen (Ed.), *Brain theory – spatio-temporal aspects of brain function* (pp. 131–45). New York: Elsevier.

Pulvermüller, F. (1994). Syntax und Hirnmechanismen. Perspektiven einer multidisziplinären Sprachwissenschaft. *Kognitionswissenschaft, 4*, 17–31.

Pulvermüller, F. (1995). Agrammatism: behavioral description and neurobiological explanation. *Journal of Cognitive Neuroscience, 7*, 165–81.

Pulvermüller, F. (1998). On the matter of rules. Past tense formation as a test-case for brain models of language. *Network: Computation in Neural Systems, 9*, R1–51.

Pulvermüller, F. (1999a). Lexical access as a brain mechanism. *Behavioral and Brain Sciences, 22*, 50–2.

Pulvermüller, F. (1999b). Words in the brain's language. *Behavioral and Brain Sciences, 22*, 253–336.

Pulvermüller, F. (2000). Syntactic circuits: how does the brain create serial order in sentences? *Brain and Language, 71*(1), 194–9.

Pulvermüller, F. (2001). Brain reflections of words and their meaning. *Trends in Cognitive Sciences, 5*, 517–24.

Pulvermüller, F. (2002). A brain perspective on language mechanisms: from discrete neuronal ensembles to serial order. *Progress in Neurobiology, 67*, 85–111.

Pulvermüller, F., Assadollahi, R., & Elbert, T. (2001). Neuromagnetic evidence for early semantic access in word recognition. *European Journal of Neuroscience, 13*(1), 201–5.

Pulvermüller, F., Birbaumer, N., Lutzenberger, W., & Mohr, B. (1997). High-frequency brain activity: its possible role in attention, perception and language processing. *Progress in Neurobiology, 52*, 427–45.

Pulvermüller, F., Eulitz, C., Pantev, C., Mohr, B., Feige, B., Lutzenberger, W., Elbert, T., & Birbaumer, N. (1996). High-frequency cortical responses reflect lexical processing: an MEG study. *Electroencephalography and Clinical Neurophysiology, 98*, 76–85.

Pulvermüller, F., Härle, M., & Hummel, F. (2000). Neurophysiological distinction of verb categories. *Neuroreport, 11*(12), 2789–93.

Pulvermüller, F., Hummel, F., & Härle, M. (2001). Walking or Talking?: Behavioral and neurophysiological correlates of action verb processing. *Brain and Language, 78*, 143–68.

Pulvermüller, F., Keil, A., & Elbert, T. (1999). High-frequency brain activity: perception or active memory? *Trends in Cognitive Sciences, 3*, 250–2.

Pulvermüller, F., Kujala, T., Shtyrov, Y., Simola. J., Tiitinen, H., Alku, P., Alho, K., Martinkauppi, S., Ilmoniemi, R. J., & Näätänen, R. (2001). Memory traces for words as revealed by the mismatch negativity. *NeuroImage, 14*, 607–16.

Pulvermüller, F., Lutzenberger, W., & Preissl, H. (1999). Nouns and verbs in the intact brain: evidence from event-related potentials and high-frequency cortical responses. *Cerebral Cortex, 9*, 498–508.

Pulvermüller, F., & Mohr, B. (1996). The concept of transcortical cell assemblies: a key to the understanding of cortical lateralization and interhemispheric interaction. *Neuroscience and Biobehavioral Reviews, 20*, 557–66.

Pulvermüller, F., Mohr, B., & Schleichert, H. (1999). Semantic or lexico-syntactic factors: what determines word-class specific activity in the human brain? *Neuroscience Letters, 275*, 81–4.

Pulvermüller, F., Mohr, B., Sedat, N., Hadler, B., & Rayman, J. (1996). Word class specific deficits in Wernicke's aphasia. *Neurocase, 2*, 203–12.

Pulvermüller, F., & Preissl, H. (1991). A cell assembly model of language. *Network: Computation in Neural Systems, 2*, 455–68.

Pulvermüller, F., Preissl, H., Lutzenberger, W., & Birbaumer, N. (1996). Brain rhythms of language: nouns versus verbs. *European Journal of Neuroscience, 8*, 937–41.

Pulvermüller, F., & Schumann, J. H. (1994). Neurobiological mechanisms of language acquisition. *Language Learning, 44*, 681–734.

Rastle, K., Davis, M. H., Marslen-Wilson, W. D., & Tyler, L. K. (2000). Morphological and semantic effects in visual word recognition: a time-course study. *Language and Cognitive Processes, 15*, 507–37.

Rauschecker, J. P., & Singer, W. (1979). Changes in the circuitry of the kitten visual cortex are gated by postsynaptic activity. *Nature, 280*, 58–60.

Redlich, A. N. (1993). Redundancy reduction as a strategy for unsupervised learning. *Neural Computation, 3*, 289–304.

Reichardt, W., & Varju, D. (1959). Übertragungseigenschaften im Auswertesystem für das Bewegungssehen. *Zeitschrift für Naturforschung, 14b*, 674–89.

Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. *Trends in Neurosciences, 21*, 188–94.

Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research, 3*, 131–141.

Rizzolatti, G., Luppino, G., & Matelli, M. (1998). The organization of the cortical motor system: new concepts. *Electroencephalography and Clinical Neurophysiology, 106*, 283–96.

Rosch, E., Mervis, C. B., Gray, W., Johnson, D., & Boyes-Bream, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8*, 382–439.

Rosenblatt, F. (1959). Two theorems of statistical separability in the perceptron. In M. Selfridge (Ed.), *Mechanisation of thought processes: proceedings of a symposium held at the National Physical Laboratory*. London: HMSO.

Rugg, M. D. (1983). Further study of the electrophysiological correlates of lexical decision. *Brain and Language, 19*, 142–52.

Rugg, M. D. (1990). Event-related potentials dissociate repetition effects of high- and low-frequency words. *Memory and Cognition, 18*, 367–79.

Rumelhart, D. E., Hinton, G., & Williams, R. (1986). Learning internal representations by backpropagation. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: explorations in the mircrostructure of cognition*. Cambridge, MA: MIT Press.

Rumelhart, D. E., & McClelland, J. L. (1986). On learning the past tense of English verbs. In J. L. McClelland & D. E. Rumelhart (Eds.), *Parallel distributed processing: explorations in the microstructure of cognition*. Cambridge, MA: MIT Press.

Rumelhart, D. E., & McClelland, J. L. (1987). Learning the past tense of English verbs: implicit rules or parallel distributed processing. In B. MacWhinney (Ed.), *Mechanisms of language acquisition*. Hillsdale, NJ: Erlbaum.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*(5294), 1926–8.

Salmelin, R., Helenius, P., & Kuukka, K. (1999). Only time can tell – words in context. *Behavioral and Brain Sciences, 22*, 300.

Scheibel, A. B., Paul, L. A., Fried, I., Forsythe, A. B., Tomiyasu, U., Wechsler, A., Kao, A., & Slotnick, J. (1985). Dendritic organization of the anterior speech area. *Experimental Neurology, 87*, 109–17.

Schnelle, H. (1996a). Approaches to computational brain theories of language – a review of recent proposals. *Theoretical Linguistics, 22*, 49–104.

Schnelle, H. (1996b). *Die Natur der Sprache. Die Dynamik der Prozesse des Sprechens und Verstehens* (2 ed.). Berlin, New York: Walter de Gruyter.

Schnelle, H. (1996c). Net-linguistic approaches to linguistic structure, brain topography, and cerebral processes.

Schumann, J. H. (1997). *The neurobiology of affect in language*. Oxford: Blackwell Publishers.

Seldon, H. L. (1985). The anatomy of speech perception. Human auditory cortex. In A. Peters & E. G. Jones (Eds.), *Cerebral cortex, Vol. 5. Association and auditory cortices* (pp. 273–327). London: Plenum Press.

Shallice, T. (1988). *From neuropsychology to mental structure*. New York: Cambridge University Press.

Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication*. Urbana: University of Illinois Press.

Shastri, L., & Ajjanagadde, V. (1993). From simple associations to systematic reasoning: a connectionist representation of rules, variables and dynamic bindings using temporal synchrony. *Behavioral and Brain Sciences, 16*, 417–94.

Shtyrov, Y., Kujala, T., Palva, S., Ilmoniemi, R. J., & Näätänen, R. (2000). Discrimination of speech and of complex nonspeech sounds of different temporal structure in the left and right cerebral hemispheres. *NeuroImage, 12*(6), 657–63.

Shtyrov, Y., & Pulvermüller, F. (2002a). Memory traces for inflectional affixes as shown by the mismatch negativity. *European Journal of Neuroscience, 15*, 1085–91.

Shtyrov, Y., & Pulvermüller, F. (2002b). Neurophysiological evidence of memory traces for words in the human brain. *Neuroreport.* In press.

Singer, W., Engel, A. K., Kreiter, A. K., Munk, M. H. J., Neuenschwander, S., & Roelfsema, P. R. (1997). Neuronal assemblies: necessity, signature and detectability. *Trends in Cognitive Sciences, 1*, 252–62.

Singer, W., & Gray, C. M. (1995). Visual feature integration and the temporal correlation hypothesis. *Annual Review in Neuroscience, 18*, 555–86.

Skrandies, W. (1998). Evoked potential correlates of semantic meaning – a brain mapping study. *Cognitive Brain Research, 6*, 173–83.

Skrandies, W. (1999). Early effects of semantic meaning on electrical brain activity. *Behavioral and Brain Sciences, 22*, 301.

Sommer, F. T., & Wennekers, T. (2000). Modelling studies on the computational function of fast temporal structure in cortical circuit activity. *Journal of Physiology, Paris, 94*, 473–88.

Sougné, J. (1998). Connectionism and the problem of multiple instantiation. *Trends in Cognitive Sciences, 2*, 183–9.

Spitzer, M., Kischka, U., Guckel, F., Bellemann, M. E., Kammer, T., Seyyedi, S., Weisbrod, M., Schwartz, A., & Brix, G. (1998). Functional magnetic resonance imaging of category-specific cortical activation: evidence for semantic maps. *Cognitive Brain Research, 6*, 309–19.

Steinmetz, H., Volkmann, J., Jancke, L., & Freund, H. J. (1991). Anatomical left–right asymmetry of language-related temporal cortex is different in left- and right-handers. *Annals of Neurology, 29*, 315–19.

Sterr, A., Muller, M. M., Elbert, T., Rockstroh, B., Pantev, C., & Taub, E. (1998). Perceptual correlates of changes in cortical representation of fingers in blind multifinger Braille readers. *Journal of Neuroscience, 18*(11), 4417–23.

Swinney, D., Onifer, W., Prather, P., & Hirshkowitz, M. (1979). Semantic fascilation across sensory modalities in the processing of individual words and sentences. *Memory and Cognition, 7*, 159–65.

Tallon-Baudry, C., & Bertrand, O. (1999). Oscillatory gamma activity in humans and its role in object representation. *Trends in Cognitive Sciences, 3*, 151–61.

Tesnière, L. (1953). *Esquisse d'une syntax structurale*. Paris: Klincksieck.

Tesnière, L. (1959). *Eléments de syntaxe structurale*. Paris: Klincksieck.

Tranel, D., & Damasio, A. R. (1999). The neurobiology of knowledge retrieval. *Behavioral and Brain Sciences, 22*, 303.

Tsumoto, T. (1992). Long-term potentiation and long-term depression in the neocortex. *Progress in Neurobiology, 39*, 209–28.

Ullman, M., Corkin, S., Coppola, M., Hickok, G., Growdon, J., Koroshetz, W., & Pinker, S. (1997). A neural dissociation within language: evidence that the mental dictionary is part of declarative memory, and that grammatical rules are processed by the procedural system. *Journal of Cognitive Neuroscience, 9*, 266–76.

Vaadia, E., Haalman, I., Abeles, M., Bergman, H., Prut, Y., Slovin, H., & Aertsen, A. (1995). Dynamics of neuronal interactions in monkey cortex in relation to behavioural events. *Nature, 373*, 515–18.

Varju, D., & Reichardt, W. (1967). Übertragungseigenschaften im Auswertesystem für das Bewegungssehen II. *Zeitschrift für Naturforschung, 22b*, 1343–51.

von der Malsburg, C. (1985). Nervous structures with dynamical links. *Bericht der Bunsengesellschaft für Physikalische Chemie, 89*, 703–10.

von der Malsburg, C. (1995). Binding in models of perception and brain function. *Current Opinions in Neurobiology, 5*, 520–26.

Waibel, A. H., Hanazawa, T., Hinton, G., Shikano, K., & Lang, K. J. (1995). Phoneme recognition using time-delay neural networks. In Y. Chauvin & D. E. Rumelhart (Eds.), *Backpropagation: theory, architectures, and applications. Developments in connectionist theory* (pp. 35–61). Hillsdale, NJ: Lawrence Erlbaum Associates.

Warburton, E., Wise, R. J. S., Price, C. J., Weiller, C., Hadar, U., Ramsay, S., & Frackowiak, R. S. J. (1996). Noun and verb retrieval by normal subjects: studies with PET. *Brain, 119*, 159–79.

Warrington, E. K., & McCarthy, R. A. (1983). Category specific access dysphasia. *Brain, 106*, 859–78.

Warrington, E. K., & McCarthy, R. A. (1987). Categories of knowledge: further fractionations and an attempted integration. *Brain, 110*, 1273–96.

Warrington, E. K., & Shallice, T. (1984). Category specific semantic impairments. *Brain, 107*, 829–54.

Weiller, C., Isensee, C., Rijntjes, M., Huber, W., Müller, S., Bier, D., Dutschka, K., Woods, R. P., North, J., & Diener, H. C. (1995). Recovery from Wernicke's aphasia: a positron emission tomography study. *Annals of Neurology, 37*, 723–32.

Wennekers, T., & Palm, G. (1996). Controlling the speed of synfire chains. *Proceedings of the International Conference on Artificial Neural Networks,* 451–6.

Wermter, S., & Elshaw, M. (2002). A neurocognitively inspired modular approach to self-organisation of action verb processing. *Connection Science,* In press.

Wernicke, C. (1874). *Der aphasische Symptomencomplex. Eine psychologische Studie auf anatomischer Basis.* Breslau: Kohn und Weigert.

Weyerts, H., Penke, M., Dohrn, U., Clahsen, H., & Münte, T. (1997). Brain potentials indicate differences between regular and irregular German plurals. *NeuroReport, 8,* 957–62.

Wickelgren, W. A. (1969). Context-sensitive coding, associative memory, and serial order in (speech) behavior. *Psychological Review, 76,* 1–15.

Wickens, J. R. (1993). *A theory of the striatum.* Oxford: Pergamon Press.

Willshaw, D. J. Buneman, O. P., & Longuet-Higgins, H. C. (1969). Non-holographic associative memory. *Nature, 222* (197), 960–2.

Willwacher, G. (1976). Faehigkeiten eines assoziativen Speichersystems im Vergleich zu Gehirnfunktionen. *Biological Cybernetics, 24*, 181–98.

Willwacher, G. (1982). Storage of a temporal pattern sequence in a network. *Biological Cybernetics, 43*, 115–26.

Winograd, T. (1983). *Language as a cognitive process, vol. 1: Syntax*. Reading, MA: Addision-Wesley.

Wise, R., Chollet, F., Hadar, U., Fiston, K., Hoffner, E., & Frackowiak, R. (1991). Distribution of cortical neural networks involved in word comprehension and word retrieval. *Brain, 114*, 1803–17.

Wittgenstein, L. (1953). *Philosophical Investigations.* Oxford: Blackwell Publishers.

Woods, B. T. (1983). Is the left hemisphere specialized for language at birth? *Trends in Neurosciences, 6*, 115–17.

Woods, W. A. (1973). An experimental parsing system for transition network grammars. In R. Rustin (Ed.), *Natural language processing* (pp. 111–154). New York: Algorithmics Press.

Young, M. P., Scannell, J. W., & Burns, G. (1995). *The analysis of cortical connectivity.* Heidelberg: Springer.

Zaidel, E. (1976). Auditory vocabulary of the right hemisphere following brain bisection or hemideocortication. *Cortex, 12*, 191–211.

Zaidel, E. (1985). Language in the right hemisphere. In D. F. Benson & E. Zaidel (Eds.), *The dual brain* (pp. 205–31). New York: Guilford.

Zaidel, E. Kasher, A., Soroker, N., Batori, G., Giora, R., & Graves, D. (2000). Hemispheric contributions to pragmatics. *Brain Cognition, 43*(1–3), 438–43.

Zaidel, E., & Rayman, J. (1994). Interhemispheric control in the normal brain: evidence from redundant bilateral presentation. In C. Umilta & M. Moscovitch (Eds.), *Attention and performance XV: concious and subconcious information processing* (pp. 477–504). Boston, MA: MIT Press.

Zatorre, R. J., Evans, A. C., Meyer, E., & Gjedde, A. (1992). Lateralization of phonetic and pitch discremination in speech processing. *Science, 256*, 846–49.

Zhou, X., & Marslen-Wilson, W. (2000). The relative time course of semantic and phonological activation in reading Chinese. *Journal of Experimental Psychology: Learning, Memory and Cognition, 26*(5), 1245–65.

Zhou, Y. D., & Fuster, J. M. (2000). Visuo-tactile cross-modal associations in cortical somatosensory cells. *Proceedings of the National Academy of Sciences, USA, 97*(17), 9777–82.

Zipser, D., Kehoe, B., Littlewort, G., & Fuster, J. (1993). A spiking network model of short-term active memory. *Journal of Neuroscience, 13*(8), 3406–20.

Zohary, E. (1992). Population coding of visual stimuli by cortical neurons tuned to more than one dimension. *Biological Cybernetics, 66*, 265–72.

Abbreviations

General

$\rightarrow$	"is rewritten as"; symbol for formulating rules of phrase structure grammars
$\longrightarrow$	"activates"; symbol for asymmetric reciprocal connection, the basic element for establishing serial order in syntactic circuits
$\leftrightarrow$	"activates"; symbol for symmetric reciprocal connection
$\Rightarrow$	"is followed by," "causes"
A, B, . . .	words or morphemes
a, b, . . .	lexical categories
$\alpha, \beta \ldots$	neuronal elements – neurons, cell assemblies, neuronal sets, or sequence sets
AB	sequence of symbols A and B
ab	sequence of symbols belonging to categories a and b
α_{pi}	i^{th} sequence set sensitive to information in the past ("past set") representing sequences of symbols that end with a symbol from category a
α_{fj}	j^{th} sequence set sensitive to information in the future ("future set") representing sequences of symbols that begin with a symbol from category a
$\mathrm{S}_p, \mathrm{S}_q, \ldots$	Neuronal sets (input or sequencing units)

Fragments of Neuronal Grammar

$\alpha \leftrightarrow \lambda$	assignment formula: the word/morpheme representation α is assigned to the lexical category representation λ
$\alpha(p_1, \ldots, p_m/*/ f_1, \ldots, f_n)$	valence formula: lexical category representation α includes m backward-looking past sets $p_1, \ldots, p_m$ and n forward-looking future sets $f_1, \ldots, f_n$.
$\alpha(f_i) \longrightarrow \beta(p_j)$	there is a directed connection from the forward-looking sequence set f_i belonging to category representation α to the backward-looking sequence set p_j of category representation β.

Syntactic and Lexical Categories

Note: Abbreviations of lexical categories are used to refer to sets of lexical items and for labeling their neuronal representations.

S	sentence
NP	noun phrase or nominal phrase
VP	verb phrase or verbal phrase
Det	determiner or article
N	noun
Prop	proper name
N1, Nnom	nominative noun
N3, Ndat	dative noun
N4, Nakk	accusative noun
V	verb
Vi	intransitive verb
Vt	transitive verb
V(t)sub	transitive verb in subordinate sentence
Vip	intransitive particle verb

Vtp	transitive particle verb (tp is sometimes omitted)
V1, V14, V134, etc.	verb subcategories specified in terms of the noun complements required (V14 = verb requiring N1 and N4)
Vp	verb particle
Vs	verb suffix
Tr	translative

Author Index

Subject Index